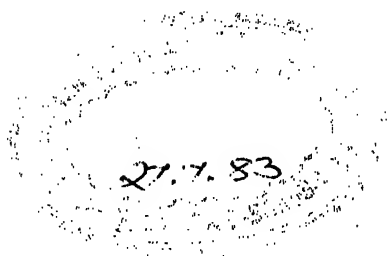


INTERNATIONAL SERIES IN PHYSICS

LEE A. DuBRIDGE, CONSULTING EDITOR



INTRODUCTION
TO MODERN PHYSICS

Donated by

Mrs. Yemuna Bappu

to

The Indian Institute of Astrophysics

from the personal collection

of

Dr. M. K. V. Bappu

*The quality of the materials used in the manufacture
of this book is governed by continued postwar shortages.*

INTERNATIONAL SERIES IN PHYSICS

LEE A. DuBRIDGE, *Consulting Editor*

- Bacher and Goudsmit*—ATOMIC ENERGY STATES
Bitter—INTRODUCTION TO FERROMAGNETISM
Clark—APPLIED X-RAYS
Condon and Morse—QUANTUM MECHANICS
Curtis—ELECTRICAL MEASUREMENTS
Davey—CRYSTAL STRUCTURE AND ITS APPLICATIONS
Edwards—ANALYTIC AND VECTOR MECHANICS
Eldridge—THE PHYSICAL BASIS OF THINGS
Hardy and Perrin—THE PRINCIPLES OF OPTICS
Harnwell—PRINCIPLES OF ELECTRICITY AND ELECTRO-
MAGNETISM
Harnwell and Livingood—EXPERIMENTAL ATOMIC PHYSICS
Houston—PRINCIPLES OF MATHEMATICAL PHYSICS
Hughes and DuBridge—PHOTOELECTRIC PHENOMENA
Hund—HIGH-FREQUENCY MEASUREMENTS
Hund—PHENOMENA IN HIGH-FREQUENCY SYSTEMS
Kemble—THE FUNDAMENTAL PRINCIPLES OF QUANTUM
MECHANICS
Kennard—KINETIC THEORY OF GASES
Koller—THE PHYSICS OF ELECTRON TUBES
Morse—VIBRATION AND SOUND
Muskat—THE FLOW OF HOMOGENEOUS FLUIDS THROUGH
POROUS MEDIA
Pauling and Goudsmit—THE STRUCTURE OF LINE SPECTRA
Richmyer and Kennard—INTRODUCTION TO MODERN
PHYSICS
Ruark and Urey—ATOMS, MOLECULES AND QUANTA
Seitz—THE MODERN THEORY OF SOLIDS
Slater—INTRODUCTION TO CHEMICAL PHYSICS
Slater—MICROWAVE TRANSMISSION
Slater and Frank—INTRODUCTION TO THEORETICAL
PHYSICS
Smythe—STATIC AND DYNAMIC ELECTRICITY
Stratton—ELECTROMAGNETIC THEORY
White—INTRODUCTION TO ATOMIC SPECTRA
Williams—MAGNETIC PHENOMENA
-

Dr. F. K. Richtmyer was consulting editor of the series from its inception in 1929 until his death in 1939.

INTRODUCTION TO MODERN PHYSICS

BY

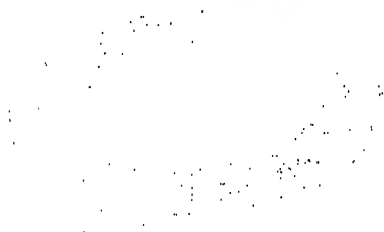
F. K. RICHTMYER

*Late Professor of Physics
at Cornell University*

AND

E. H. KENNARD

*Professor of Physics
at Cornell University*



THIRD EDITION
SEVENTH IMPRESSION

McGRAW-HILL BOOK COMPANY, INC
NEW YORK AND LONDON
1942

INTRODUCTION TO MODERN PHYSICS

COPYRIGHT, 1928, 1934, 1942, BY THE
MCGRAW-HILL BOOK COMPANY, INC.

PRINTED IN THE UNITED STATES OF AMERICA

*All rights reserved. This book, or
parts thereof, may not be reproduced
in any form without permission of
the publishers.*

PREFACE TO THE THIRD EDITION

After the sudden death of Professor Richtmyer, it was decided that the writer should undertake the preparation of a new edition of his "Introduction to Modern Physics." Closer inspection showed that something like a major operation would be necessary, so greatly has the scene changed in physical research during the last ten years. In atomic theory, wave mechanics has become thoroughly established; and experimental advance has been rapid in the physics of the nucleus and of fundamental particles, as observed in the phenomena of cosmic rays. If the book was to remain true to its title, extensive additions seemed necessary, and to make room for them the existing material had to be shortened; at the same time, much of the discussion needed recasting in order to give due recognition to the triumph of wave mechanics.

In making these changes, the writer has striven to retain as much as possible of the characteristic features of the book. Nevertheless, its friends will no doubt be shocked to discover that only a minor part of it stands substantially as in the second edition. The double process of shortening and of modernizing the viewpoint was found to necessitate extensive rearrangement and much rewriting of the material. The writer can only hope that he has not fallen too far short of his goal in attempting to produce an up-to-date equivalent of the first edition.

Briefly, the course of the revision has been as follows. The historical introduction, abbreviated, is followed by a single chapter, rewritten, on those topics in electromagnetism which are needed for the subsequent discussion and are not always adequately treated in general textbooks. Between the chapters on the photoelectric effect and on the origin of the quantum theory is inserted a short chapter on relativity. Then comes a single chapter, replacing Chaps. IX and X, containing the essential ideas concerning the nuclear atom, spectral series, and atomic quantum states. The Bohr theory of hydrogen is retained, because of its pictorial value, but with a clear statement as to its true status. A descriptive chapter on wave mechanics is then followed by a single chapter, replacing Chaps. XI and XII, on the theory of the periodic table and on optical spectroscopy. The chapter on specific heats, which might perhaps have been omitted, is inserted next. The chapter on X-rays has been thoroughly revised

in collaboration with Prof. L. G. Parratt. The book closes with the chapter on the nucleus, considerably extended, and a new chapter on cosmic-ray phenomena.

The writer takes pleasure in acknowledging his indebtedness to many of his colleagues in Cornell University, in particular to Prof. L. G. Parratt, to Prof. B. Rossi for assistance in the field of cosmic rays, and to Dr. C. W. Gartlein for photographic copies of illustrations. Grateful acknowledgment is also made for permission to use figures from other sources: to G. Herzberg and to Prentice-Hall, Inc., for Figs. 58, 59, 129, 132, 133; to A. H. Compton and S. K. Allison and to D. Van Nostrand Company, Inc., for Figs. 169, 170; to H. E. White, for Figs. 98, 118, 122; to Prof. C. D. Anderson for photographs for Figs. 226, 227; and to Prof. L. G. Parratt for a cut for Fig. 180. Furthermore, in accord with the statement in the preface to the second edition, grateful acknowledgment is made to Dr. K. T. Bainbridge for Fig. 187; to Dr. C. J. Davisson and L. H. Germer (and to the editors of the *Bell System Technical Journal* and of the *Proceedings of the National Academy of Sciences*) for Figs. 79 to 84; to Prof. G. P. Thomson (and Cornell University) for Fig. 86(b); to Dr. E. Rupp (and the editors of *Annalen der Physik*) for Fig. 86(c); and to Dr. I. Fankuchen for making the plate from which Fig. 86(a) was reproduced. Finally, the writer is greatly indebted to Dr. W. M. Cady and to other users of the book for invaluable criticisms and suggestions.

E. H. KENNARD.

ITHACA, N. Y.,
January, 1942.

PREFACE TO THE FIRST EDITION

For several years, the author has given at Cornell University, and, occasionally, in summer sessions, elsewhere, a course of lectures under the title "Introduction to Modern Physical Theories." These lectures have been adapted, as far as possible, to meet the needs of two groups of students: (1) those special students in physics who, before entering the specialized graduate courses, desire a survey of the origin and development of modern physics in order the better to understand the interrelations of the more advanced courses; and (2) those students who, pursuing either academic or professional curricula and having had the usual elementary undergraduate courses in physics, wish a further bird's-eye view of the whole subject. This book is based upon these lectures and has been prepared, although rather reluctantly, as a result of the importunities of former students and other friends.

The purpose of this book is, frankly, pedagogical. The author has attempted to present such a discussion of the origin, development, and present status of *some* of the more important concepts of physics, *classical as well as modern*, as will give to the student a correct perspective of the growth and present trend of physics as a whole. Such a perspective is a necessary basis—so the author, at least, believes—for a more intensive study of any of the various subdivisions of the subject. While for the student whose interests are cultural, or who is to enter any of the professions directly or indirectly related to physics, such as engineering, chemistry, astronomy, or mathematics, an account of modern physics which gives the *origin* of current theories is likely to be quite as interesting and valuable as is a categorical statement of the theories themselves. Indeed, in *all* branches of human knowledge the "why" is an absolutely indispensable accompaniment to the "what." "Why?" is the proverbial question of childhood. "Why?" inquires the *thoughtful* (!) student in classroom or lecture hall. "Why?" demands the venerable scientist when listening to an exposition of views held by a colleague. Accordingly, if this book seems to lay somewhat greater emphasis on matters which are frequently regarded as historical, or, if here and there a classical experiment is described in greater detail than is customary, it is with a desire to recognize the importance of "why."

If one were to attempt to answer all of the "why's" raised by an intelligent auditor in answer to a categorical statement, such as, "The atom of oxygen is composed of eight electrons surrounding a nucleus containing four alpha particles," one would have to expound a large part of physical science from Copernicus to Rutherford and Bohr. To attempt a statement of even the more important concepts, hypotheses, and laws of modern physics and of their origin and development would be an encyclopedic task which, at least in so far as concerns the aim of this book, would fall of its own weight. Accordingly, it has been necessary to select those parts of the subject which best serve our purpose. This selection, as well as the method of presentation, has been based upon the experience gained in giving the above-mentioned lectures to numerous groups of students. Many very important developments, particularly the more recent ones, either have been omitted entirely or have been given only a passing comment. And even in those parts of the subject which have been discussed, there has been no attempt to bring the discussion strictly up to date. Indeed, with the present rapid growth of physics, it would be quite impossible for any *book*, even a special treatise, to be strictly up to date. Happily, for our purpose, up-to-date-ness is not an imperative requisite, since it is assumed that the student who wishes the *latest* knowledge will consult the current periodicals.

In this connection, it should be emphasized that this book is an *introduction* to modern physical theories and is intended neither as a compendium of information nor as a critical account of any of the subjects discussed. In preparing the manuscript, the author has consulted freely the many very excellent texts which deal with the various special topics. Save for here and there a very minor item, or an occasional novelty in presentation, the book makes no claim to originality, except, perhaps, as regards the viewpoint from which some parts have been written.

It is assumed that the student is familiar with the elementary principles of calculus, for no account of modern physics can dispense with at least a limited amount of mathematical discussion, if for no other reason than to emphasize the fact that, in the progress of physics, *theory* and *experiment* have gone hand in hand. Partly, however, for the sake of brevity and partly in the attempt always to keep the underlying physical principles in the foreground, numerous "short-cuts" and simplifications, some of them perhaps rather questionable from a precise standpoint, have been introduced. These elisions should cause no confusion.

The student who, in his educational career, has reached the point where he can, with profit, pursue a course based on such a book as this, has passed beyond the stage where he assimilates only the material presented in lecture or class and has come to regard a "course" as a channel to guide his own independent studies, branching out from the "course" in such directions as his fancy or interests may lead him. It is hoped that students reading this book will do likewise. Deliberately, the author has not given a collected bibliography at the end of each chapter, or a list of problems and suggested topics for study. Rather, references, in most cases to original sources, have been given at appropriate points in the text, and it is hoped that, starting from these references, the student will prepare his own bibliography of such parts of the subject as appeal to him. The advantage to the student of such a procedure is obvious. Quite apart from the value of the experience gained in making contact with, and in studying, the literature of any subject, the reading of *first-hand* accounts of at least some of the more important developments will give the student a better understanding of the subject than can, in general, be gained by textbook study only. Accordingly, he will find here and there throughout this book suggestions of important articles which should be read in the original. Likewise, in many places the discussion has, of necessity, been brief, and the student is referred to special treatises for further details. Various supplementary questions and problems will also arise at numerous points as the student reads the text.

There is no more fascinating story than an account of the development of physical science as a whole. (*Any* scientist would probably make the same statement about *his own* science!) Such a study leads to certain broad generalizations which are of outstanding importance in evaluating current theories and concepts. For example, one finds that, taken by and large, the evolution of physics has been characterized by *continuity*. That is to say: With few exceptions, the ideas, concepts, and laws of physics have evolved *gradually*; only here and there do we find *outstanding* discontinuities. The discovery of photoelectricity, of X-rays, and of radioactivity represent such discontinuities and are correctly designated "discoveries." But we must use "discover" in a quite different sense when we say that J. J. Thomson "discovered" the electron. The history of the electron goes back at least to Faraday. Thomson's experiments are all the more brilliant because he succeeded in demonstrating, by direct experiment, the existence of something, evidence concerning which had been previously *indirect*. Then, there are the respective roles played by qualitative

and by quantitative work. Numerous important discoveries have been made "by investigating the next decimal place." Witness the discovery of argon. And ever since Kepler proved that the orbits of the planets are ellipses, relations expressible in *quantitative* form have carried greater weight than those which could be stated only qualitatively. For example, Rumford's experiments on the production of heat by mechanical means were suggestive. But Joule's measurement of the mechanical equivalent of heat was *convincing*. If, directly, or indirectly by inference, the author has succeeded here and there in the text in pointing out such generalizations as these, one more object which he has had in mind will have been accomplished.

The author wishes to take this occasion to acknowledge his obligations to those who have aided in the preparation of this book: to his wife, for assistance in preparing the manuscript and in proof reading; and to his many students, whose generous approbation of the lecture courses upon which the book is based has, in a large part, inspired its preparation. He is particularly indebted to Dr. J. A. Becker, of the Bell Telephone Laboratories, Inc., for his invaluable aid in reading the manuscript, pointing out numerous errors, and suggesting important improvements.

F. K. RICHTMYER.

ITHACA, N. Y.,
July, 1928.

CONTENTS

(Page numbers are given after the section headings.)

	PAGE
PREFACE TO THE THIRD EDITION	v
PREFACE TO THE FIRST EDITION	vii
INTRODUCTION	1

CHAPTER I

HISTORICAL SKETCH	5
-----------------------------	---

First Period: Earliest Times to A. D. 1550: 1. The Greeks (5) 2. Thales of Miletus (5) 3. Pythagoras (5) 4. Anaxagoras and Empedocles (5) 5. Democritus (6) 6. Aristotle (6) 7. Aristarchus (9) 8. Archimedes (9) 9. From the Greeks to Copernicus (9) 10. The Copernican System (11).

Second Period (A. D. 1550-1800): Rise of the Experimental Method. 11. Galileo Galilei (12) 12. Tycho Brahe and Kepler (17) 13. The Experimental Method Spreads (19) 14. Sir Isaac Newton (21) 15. Newton's Contemporaries (28) 16. Mechanics during the Eighteenth Century (28) 17. Heat during the Eighteenth Century (29) 18. Light during the Eighteenth Century (29) 19. Electricity during the Eighteenth Century (30) 20. Close of the Second Period (32).

Third Period (A. D. 1800-1890): The Rise of Classical Physics. 21. The Nineteenth Century in Physics (32) 22. Heat and Energy (33) 23. Light (34) 24. Electricity and Magnetism (36) 25. Michael Faraday (38) 26. Joseph Henry (45) 27. James Clerk Maxwell (46) 28. The Completion of Electromagnetic Theory (50).

CHAPTER II

ELECTROMAGNETIC WAVES AND MOVING CHARGES	51
--	----

29. Displacement Currents (51) 30. Maxwell's Equations (54) 31. Energy and Momentum in the Electromagnetic Field (59) 32. Electromagnetic Waves (61) 33. Velocity of Electromagnetic Waves (63) 34. Energy of Electromagnetic Waves (65) 35. Momentum of Electromagnetic Waves. Radiation Pressure (68) 36. Field of a Uniformly Moving Point Charge (70) 37. Radiation Field of an Accelerated Point Charge (73) 38. Energy Radiated by an Accelerated Point Charge (76) 39. Electromagnetic Mass (80).

CHAPTER III

THE PHOTOELECTRIC AND THERMIONIC EFFECTS	83
--	----

40. Discovery of the Photoelectric Effect (83) 41. A Problem (85) 42. Electricity in Matter (86) 43. The Zeeman Effect (87) 44. The

Discovery of the Electron (92) 45. Electronic Magnitudes (95) 46. Photoelectrons (97) 47. Relation between Photoelectric Current and Intensity of Illumination of the Cathode (100) 48. Energy Distribution of Photoelectrons (100) 49. Relation between the Velocities of Photoelectrons and the Frequency of the Light (103) 50. Other Properties of Photoelectric Emission (104) 51. Source of the Photoelectric Energy (105) 52. The Photoelectric Effect and the Corpuscular Theory of Light (108) 53. Thermionic Emission (109) 54. Relation between Thermionic and Photoelectric Constants (111) 55. Velocities of Emission of Thermions (112) 56. Theories of Electrons in Metals (113) 57. Origin of Photoelectrons and Thermions (117).

CHAPTER IV

THE THEORY OF RELATIVITY. 122

58. Newtonian Relativity (122) 59. Relativity and the Propagation of Light (124) 60. The Michelson-Morley Experiment (126) 61. The New Relativity of Einstein (130) 62. Simultaneity and Time Order (132) 63. The Lorentz Transformation (134) 64. Contractions in Space and Time (136) 65. The Transformation of Velocities (138) 66. General Significance of the Theory of Relativity (139) 67. Relativistic Mechanics. The Variation of Mass (140) 68. Force and Kinetic Energy (143) 69. A Relation between Mass and Energy (145) 70. Relativity and Electromagnetism (147) 71. General Theory of Relativity (150) 72. Einstein's Law of Gravitation (152).

CHAPTER V

THE ORIGIN OF THE QUANTUM THEORY 154

73. Thermal Radiation (154) 74. Some Fundamental Concepts and Definitions (155) 75. The Black Body and the Isothermal Enclosure (160) 76. Relation between Emissive Power and Absorptivity (162) 77. Pressure due to Isotropic Radiation (164) 78. Deduction of the Stefan-Boltzmann Law (166) 79. Experimental Verification of the Stefan-Boltzmann Law (169) 80. Reflection from a Moving Mirror (170) 81. Effect of an Adiabatic Expansion upon Black-body Radiation (173) 82. The Wien Displacement Law (175) 83. The Formula for Black-body Radiation (178) 84. Degrees of Freedom (179) 85. Relation between Energy per Degree of Freedom and the Temperature (180) 86. Verification of Equipartition of Energy in the Brownian Motion (183) 87. Degrees of Freedom in an Enclosure (184) 88. The Rayleigh-Jeans Formula (189) 89. Other Formulas for Black-body Radiation (190) 90. Distribution of Harmonic Oscillators in Thermal Equilibrium (191) 91. Average Energy of Oscillators in Equilibrium (194) 92. Planck's Quantum Hypothesis (197) 93. Planck's Radiation Law (200).

CHAPTER VI

THE NUCLEAR ATOM AND THE ORIGIN OF SPECTRAL LINES. 206

94. Spectroscopic Units (207) 95. Early Search for Regularities in Spectra (209) 96. Spectral Series and their Interrelations (212) 97. Further Relationships between Series. Spectral Terms (215) 98. Early

Views on Atomic Structure (217)	99. The Thomson Atom (219)	100.
The Scattering of Alpha Particles by Atoms (220)	101. Rutherford's Nuclear Atom (222)	102.
The Bohr Theory of Atomic Hydrogen (226)	103. Quantum States of One Electron in an Atom (230)	104.
Spectrum of a One-electron Atom (233)	105. The Spectrum of Atomic Hydrogen. Energy Levels and Spectral Series (237)	106.
Ionized Helium (239)	107. Energy Levels and Series Relationships for Sodium (241)	108.
Excitation and Ionization of Atoms by Electrons (243)	109. Absorption and Re-emission of Radiation (247)	110.
The Boltzmann Distribution Law (255)	111. The Extension of Bohr's Theory (256).	

CHAPTER VII

WAVE MECHANICS	259
--------------------------	-----

112. Matter Waves (259)	113. Mechanics as Geometrical Optics of the Waves (262)	114. Refraction of Matter Waves (263)	115. Fermat's Principle and the Principle of Least Action (266)	116. The de Broglie Wave Length (269)	117. Experiments on Electron Waves (273)	118. Diffraction of Molecule Waves (283)	119. Schrödinger's Wave Equation (284)	120. Physical Significance of ψ (289)	121. The Indeterminacy Principle (292)	122. Stationary States (294)	123. The Harmonic Oscillator (296)	124. The One-electron Atom (297)	125. Emission and Absorption of Radiation (304)	126. Relativistic Effects and Electron Spin (309).
-------------------------	---	---------------------------------------	---	---------------------------------------	--	--	--	--	--	------------------------------	------------------------------------	----------------------------------	---	--

CHAPTER VIII

ATOMIC STRUCTURE AND OPTICAL SPECTRA.	311
---	-----

Wave Mechanics of Many-electron Systems: 127. Wave-mechanical Fundamentals (311) 128. The Zero-order Approximation: Spin Energy Omitted (314) 129. The More Accurate Theory (319).

The Periodic Table of the Elements: 130. Atomic Numbers (320) 131. Some Features of the Periodic Table (321) 132. The Static Atom (324) 133. Wave Mechanics and the Periodic Table: the First Two Periods (325) 134. Valence Bonds (329) 135. The Third Period of the Periodic Table (332) 136. The Fourth and Fifth Periods (333) 137. The Periodic Table Concluded (335).

Optical Spectra: 138. Atomic and Molecular Spectra (337) 139. Angular Momentum in Wave Mechanics (339) 140. Total Angular Momentum and its Selection Rules (340) 141. Orbital and Spin Angular Momenta (342) 142. Alkali-type Spectra (345) 143. The Term Energies of the Alkali Metals (350) 144. The Addition of Orbital and Spin Momenta (353) 145. Number of Levels in an LS Multiplet (355) 146. Spin-orbit Interaction and Multiplet Spacing (358) 147. Fine Structure in Alkali-type Spectra (362) 148. Multiplet Levels for One-electron Atoms (366) 149. Fine Structure of Spectral Lines from One-electron Atoms (369) 150. Two-electron Spectra (372) 151. The Spectrum of Mercury (377) 152. Equivalent Electrons (379) 153. " jj " Coupling (380) 154. Spectra due to Three or More Electrons (383) 155. The Effect of a Magnetic Field on an Atom (384) 156. Classical Theory of the Magnetic Energy (387) 157. Zeeman Effect in a Huge Field (391) 158. Zeeman Effect in a Weak Field (396) 159. Zeeman

Patterns of *LS* Multiplets in a Weak Field (400) 160. The Paschen-Back Effect (403) 161. The Stark Effect (406) 162. The Stern-Gerlach Experiment (407) 163. Isotope Structure and Hyperfine Structure (409) 164. The Breadth of Spectral Lines (414) 165. Molecular Spectra (417) 166. Rotation Spectra (419) 167. Vibration-rotation Spectra (421) 168. General Theory of Molecular Quantum States (429) 169. Electronic Bands (433) 170. The Raman Effect (437).

CHAPTER IX

THE QUANTUM THEORY OF SPECIFIC HEATS. 442

171. Variation of Specific Heats of Solids with Temperature (442) 172. Classical Theory of the Specific Heats of Solids (444) 173. Einstein's Theory of the Atomic Heats of Solids (445) 174. Characteristic Frequencies (447) 175. Debye's Theory of Atomic Heats (450) 176. Experimental Test of Debye's Equation (453) 177. Molecular Heats of Mixed Solids (456) 178. The Molecular Heat of Gases: Classical Theory (457) 179. Quantum Theory of the Specific Heat of Gases (460) 180. Comparison of the Theory with Observed Specific Heats (465).

CHAPTER X

X-RAYS. 469

Early, Mostly Qualitative Developments in X-rays (1895-1912): 181. Roentgen's Discovery (469) 182. Production and Measurement of X-rays (471) 183. Classical Pulse Theory of X-rays (473) 184. The Scattering of X-rays (475) 185. Absorption and Fluorescence (482).

X-ray Spectra: 186. The Crystal Diffraction Grating (486) 187. The X-ray Spectrometer (491) 188. Bragg's Discovery of Monochromatic Characteristic Radiations (495) 189. Moseley's Law (496) 190. The Origin of X-ray Lines (498) 191. X-ray Energy Levels (504) 192. The Quantum Theory of X-ray Terms and Lines (507) 193. The Continuous X-ray Spectrum (512) 194. Intensity of the Continuous Spectrum (517).

Interactions of X-rays with Atoms: 195. The Absorption of X-rays (519) 196. The Photoelectric Effect for X-rays (523) 197. The Scattering of X-rays (528) 198. The Compton Scattering of X-rays (531) 199. The Refraction of X-rays (539) 200. Measurement of X-ray Wave Lengths by a Ruled Grating (544).

Some Recent Developments in X-ray Spectroscopy: 201. Multiple Ionization of Inner Electron Shells (545) 202. X-ray Spectra and the Outer Part of the Atom (550) 203. X-ray Spectroscopy of Solids (553).

CHAPTER XI

THE NUCLEUS. 556

The Masses of Atoms: 204. Positive Rays (556) 205. The Mass Spectrograph (559) 206. Isotopes (563).

Radioactivity: 207. Becquerel's Discovery of Radioactivity (566) 208. The Radioactive Radiations (567) 209. The Alpha-rays (568) 210. Radioactive Transformations (574) 211. Gamma-ray Spectra and Nuclear Energy Levels (577) 212. Beta-ray Spectra (582) 213. Observations on Individual Charged Particles (583).

CONTENTS

XV

PAGE

The Structure of Nuclei: 214. The Building up of Nuclei (586) 215. The General Theory of Nuclear Binding (591) 216. Nuclear Forces and the Ratio of Mass Number to Atomic Number (593) 217. The Explanation of Natural Radioactivity (597).

Artificial Transmutation of Nuclei: 218. Artificial Transmutation by Alpha Particles (599) 219. Artificial Transmutation by High-velocity Protons (602) 220. Discovery of the Neutron (606) 221. Positrons (610) 222. The Discovery of Induced Radioactivity (613) 223. Experiments with Neutrons (615) 224. The Production of High-speed Ions (621) 225. Some Typical Nuclear Reactions (625) 226. Fission of the Nucleus (638).

CHAPTER XII

COSMIC RAYS 643

227. Early Work on Cosmic Rays (643) 228. The Measurement of Cosmic-ray Ionization (646) 229. The Altitude-depth Curve (648) 230. Discovery of the Latitude Effect (650) 231. Motion of a Charged Particle in the Equatorial Plane of a Magnetic Dipole (651) 232. Motion of Charged Particles in the Magnetic Field of the Earth (656) 233. The Variation of Cosmic-ray Intensity with Latitude (661) 234. Observations on Single Cosmic-ray Particles (665) 235. Energy, Mass, and Specific Ionization of Charged Particles (669) 236. Showers and Bursts (674) 237. Theory of the Shower Phenomenon (679) 238. Mesotrons (689) 239. Origin and Fate of Mesotrons (693) 240. Conclusion (696).

APPENDIXES

I. ISOTOPIC CONSTITUTION OF THE ELEMENTS	699
II. THE PERIODIC TABLE.	704
III. FIRST IONIZATION POTENTIALS (WHERE KNOWN), LOWEST SPECTRAL TERMS, AND ELECTRON CONFIGURATION OF THE ELEMENTS	705
SOME USEFUL CONSTANTS AND RELATIONS.	708
INDEX.	709

INTRODUCTION TO MODERN PHYSICS

INTRODUCTION

The term "modern physics," taken literally, means, of course, the *sum total* of knowledge included under the head of present-day physics. In this sense, the physics of 1890 is still modern; very few statements made in a good physics text of 1890 would need to be deleted today as untrue. The principal changes required would be in a few generalizations, perhaps, to which exceptions have since been discovered, and in certain speculative theories, such as that concerning the ether, which any good physicist of 1890 would have recognized to be open to possible doubt.

On the other hand, since 1890, there have been enormous advances in physics, and some of these advances have brought into question, or have directly contradicted, certain theories that had seemed to be strongly supported by the experimental evidence.

For example, few, if any, physicists in 1890 questioned the wave theory of light. Its triumphs over the old corpuscular theory seemed to be final and complete, particularly after the brilliant experiments of Hertz, in 1887, which demonstrated, beyond doubt, the fundamental soundness of Maxwell's electromagnetic theory of light. And yet, by an irony of fate which makes the story of modern physics full of the most interesting and dramatic situations, these very experiments of Hertz brought to light a new phenomenon—the photoelectric effect—which played an important part in establishing the quantum theory. The latter theory, in many of its aspects, is diametrically opposed to the wave theory of light; indeed, the reconciliation of these two theories, each based on incontrovertible evidence, was one of the great problems of the first quarter of the twentieth century.

It will be the purpose of the following pages to give an outline of the origin, development, and present status of these parts of physics which have developed during recent decades.

But a history of the United States cannot begin abruptly with July 4, 1776. In like manner, if we would understand the full meaning of the growth of physics since 1900, we must have clearly in mind at least the main events in the development of the subject up to that time.

Accordingly, we shall begin our study by a brief account of the history of physics up to a half-century ago.

In presenting this brief historical survey, a further purpose has been kept in mind toward which it is hoped that the reader will be, ultimately at least, sympathetic. Modern scientists, with few exceptions, have grossly neglected to cultivate the history of their respective sciences. How many physicists can answer the questions: When was the law of the conservation of energy first enunciated? Who was Count Rumford? Did the concept of universal gravitation spring full grown from the head of that genius, Newton? Indeed, when did Newton live?

Just as any good American should know the essential outline of the history of his country, so any good physicist should know the principal facts in the history of physics. For, in that history, in the lives of those men whose labors have given us our subject, and in the part that physics has played in molding human thought and in contributing to modern civilization, the student will find a story which is as full of human interest and inspiration as any subject of the curriculum. What can be more inspiring than the life of Michael Faraday and his whole-souled devotion to his work? The physicist owes it to his science to possess such a knowledge of the history of physics as gives *him* a correct perspective of the development and present-day importance of the subject and, in turn, enables him to acquaint his contemporaries in other fields with these essential facts.

It is hoped, therefore, that the student who proposes to follow physics as a profession, as well as the student whose interest is largely cultural, will extend the following all too brief historical sketch by independent study, particularly of biography.

In order to make it easier to keep the essential facts in mind, we may somewhat arbitrarily divide the history of physics into four periods.

The **FIRST PERIOD** extends from the earliest times up to about 1550 A.D., this date marking approximately the beginning of the experimental method. During this long era, there was some advance in the accumulation of the *facts* of physics as a result of the observation of natural phenomena, but the development of physical *theories* was rendered impossible, partly by the speculative, metaphysical nature of the reasoning employed, but more particularly by the almost complete absence of experiment to test the correctness of such theories as were proposed. The main characteristic of this period, therefore, is *the absence of systematic experiment*.

The **SECOND PERIOD** extends from 1550 to 1800 A.D. Although numerous basic advances were made during this period—by such men

as Gilbert, Galileo, Newton, Huygens, Boyle—its most important characteristic is *the development and the firm establishment of the experimental method* as a recognized and accepted means of scientific inquiry. This period was inaugurated by the classic work of Galileo (1564–1642); but it took nearly two centuries more to overcome prejudice, dogma, and religious intolerance and to bring universal recognition, even among scientific men, of the basic principle that

... science can advance only so far as theories, themselves based upon experiment, are accepted or rejected according as they either agree with or are contrary to other experiments devised to check the theory.

The THIRD PERIOD, 1800–1890, is characterized by the development of what is now called “classical physics” in contrast with the “quantum physics” of the present century. The experiments of Count Rumford and Joule led to our present kinetic theory of heat. The observations of Thomas Young (1802) and his proposal of the principle of interference (of two beams of light) resulted ultimately in the triumph of Huygens’s wave theory of light over the corpuscular theory. And the researches of Faraday and others gave Maxwell the material for the crowning achievement of this period, the electromagnetic theory.

So profound were these and other developments that, by 1880, not a few physicists of note believed that all the important laws of physics had been discovered and that, henceforth, research would be concerned with clearing up minor problems and, particularly, with improvements of methods of measurement so as “to investigate the next decimal place.” They could not foresee that the world of physics was on the eve of a series of epoch-making discoveries destined, on the one hand, to stimulate research as never before and, on the other, to usher in an era of the application of physics to industry on a scale previously unknown.

The FOURTH PERIOD may be said to begin with the discovery of the photoelectric effect (1887). In the first decade of this period there were discovered, in rapid succession: X-rays, in 1895; radioactivity, in 1896; the electron, in 1897. The beginnings of the quantum theory date from 1900. From 1900 to 1925 the older form of the quantum theory grew to occupy a commanding position in almost every field of physics; the nuclear type of atom and its relation to the emission and absorption of radiation were developed to a high degree; research in physics, stimulated in part by these outstanding discoveries and in part by extensive industrial applications, increased to almost a new order of magnitude. Then, when physicists were

just beginning to grow accustomed to the use of either the classical theory or the quantum theory, according to the problem in hand, the theoretical papers of de Broglie, Heisenberg and Schrödinger, and the experimental work of Davisson and Germer and of G. P. Thomson, beginning about 1925, initiated the new form of quantum theory known as wave mechanics. This new theory has effected, in a radical and remarkable way, a synthesis of classical and quantum physics and has already exerted almost as profound an influence on physics as did the discoveries of Newton over two and a half centuries ago. Perhaps the historian of a few decades hence may designate 1925 as marking the beginning of a *fifth* period in the history of physics.

It is obviously far beyond the scope of this book to give a detailed account of the history of physics during each of these periods. Instead, in the first chapter we shall discuss briefly the history of physics up to the establishment of the electromagnetic theory of light. Thereafter we shall not adhere strictly to the historical method. In the first chapter, furthermore, we shall make no attempt to give a complete outline of the history of physics, but we shall select such material for presentation as best illustrates general trends and viewpoints. No mention at all will be made of the work of many distinguished physicists of the past. It is hoped that the student will fill in the gaps by supplementary reading.

The real introduction to *modern physics* begins with Chap. III. It should be emphasized again that this book is an *introduction* to the subject and is in no sense a compendium. If little or no mention is made of such important branches of physics as thermionics, crystal structure, or the theory of the solid state, it is because of the feeling that the student's interests will be served best by discussing relatively few subjects more thoroughly.

It is desirable, of course, that a book dealing with modern physics should be up to date. The subject is growing at such an astonishing rate, however, that any book, even if up to date when published, would be out of date within a few months. Accordingly, those topics have been selected for discussion which seem most likely to remain permanently important and to provide the student with the most effective starting point for further study.

CHAPTER I

HISTORICAL SKETCH

FIRST PERIOD: EARLIEST TIMES TO 1550 A.D.

1. The Greeks.—Relatively speaking, the contributions made by the Greeks to the natural sciences were far less than their contributions to mathematics, literature, art, and metaphysics. Nevertheless, in spite of their vague and misty philosophizing concerning natural phenomena and in spite of their general failure to test theory by experiment, the Greeks gave to the world much of the physics that was known up to 1400 A.D. In their writings, one finds, here and there, the germ of such fundamental modern principles as the conservation of matter, inertia, atomic theory, the finite velocity of light, and the like.

2. Thales of Miletus (624–547 B.C.), according to Aristotle, was acquainted with the attractive power of magnets and of rubbed amber. He is often said to have discovered the inclination of the ecliptic and the spherical shape of the earth,¹ but Aristotle credited him with the doctrine that the earth was cylindrical in shape and rested on water.

3. Pythagoras (580–500 B.C.) was one of the greatest of the early Greek philosophers and the founder of the Pythagorean school. He held that the earth is spherical, although the basis of this belief is not known. According to Heath,² his argument was probably a "mathematico-esthetic" one based on the idea that "the sphere is the most perfect of all figures." Pythagoras himself, and probably his immediate successors among the Pythagoreans, believed that the entire universe was spherical in shape with the earth at its center and that the sun, stars, and planets moved in independent circles around the earth as a center.

4. Anaxagoras (500–428 B.C.) and **Empedocles** (484–424 B.C.). According to Plato, Anaxagoras neglected his possessions in order to devote himself to science. He is credited with the view that the moon does not shine by its own light but that "the sun places the brightness in the moon" and "the moon is eclipsed through the interposition of the earth." Also, "The sun is eclipsed at new moon through the

¹ ROSENBERGER, F., "Geschichte der Physik," vol. I, p. 6.

² HEATH, T. L., "Aristarchus of Samos."

interposition of the moon."¹ Apparently, however, human nature, as well as human curiosity, has not changed for 2,500 years, for Anaxagoras was accused of impiety because he taught that the sun was a red-hot stone and that the moon was simply earth, and for holding this doctrine he was banished from Athens.

To Anaxagoras is due the germ of the idea of the atomic hypothesis of Democritus, who lived in the next generation. Anaxagoras denied the contention of the earlier Greeks regarding the creation or destruction of matter. He taught that changes in matter are due to combinations or separations of small, invisible particles (*spermata*). These particles themselves were conceived to be unchangeable and imperishable and different from each other in form, color, and taste. This doctrine foreshadowed the law of the conservation of matter.

Empedocles, on the other hand, reduced the elements to four—earth, water, air, and fire—through combinations and separations of which the All exists. He also held definite views concerning the phenomena of light. According to him, light is due to the emission by the luminous or visible body of small particles that enter the eye and are then returned from the eye to the body, the two "streams" giving rise to the sense of form, color, etc.

According to Aristotle, Empedocles believed that light "takes time to travel from one point to another." This idea was rejected by Aristotle, who stated that "though a movement of light might elude our observation within a short distance, that it should do so all the way from east to west is too much to assume."²

5. Democritus (460-370 B.C.) gave more definite form to the atomic hypothesis of Anaxagoras by postulating that *the universe consists of empty space and an (almost) infinite number of indivisible and invisible particles* which differ from each other in *form, position, and arrangement*. In support of this hypothesis, Democritus argues that the creation of matter is impossible, since *nothing* can come from *nothing* and, further, nothing which is can cease to exist. Aristotle³ puts the argument in this form: "If, then, some one of the things which are is constantly disappearing, why has not the whole of what is been used up long ago and vanished away?" But Aristotle rejects the atomic hypothesis which, indeed, on the basis of speculative reasoning alone, could not evolve beyond the point where Democritus left it.

6. Aristotle (384-322 B.C.), a pupil of the philosopher Plato, contributed so much to all branches of knowledge—logic, rhetoric, ethics,

¹ *Ibid.*, p. 72. Quotations cited from later Greek writers.

² Quoted by Heath, *op. cit.*, p. 93.

³ "De Generatione et Corruptione," Book I, Cap. III, translated by H. H. Joachim, the Clarendon Press Oxford, 1922.

metaphysics, psychology, natural science—that it is difficult to sift out that which is germane to a brief history of physics. Perhaps the most important single fact is the tremendous influence which, as a result of his intellectual brilliance and achievements in *many* branches of learning, he exerted for many succeeding centuries in *all* branches, physics included. Viewed from our twentieth-century vantage point, however, not a little of his reasoning concerning the physical universe sounds like piffle. For example, in “De Generatione et Corruptione,” he discusses the “coming-to-be” and the “passing-away” of things, and argues for the indestructibility of matter by saying that (Book II, Cap. X) “it is far more reasonable (to assume) that *what is* should cause the coming-to-be of *what is not* than that *what is not* should cause the being of *what is*,” which is understandable. But then follows the curious argument: “Now that which is being moved *is*, but that which is coming to be *is not*; hence, also, motion is prior to coming-to-be . . . and we assert that motion causes *coming-to-be*.” But *coming-to-be* and *passing-away* are two processes contrary to one another. Therefore, says Aristotle, we must look for *two motions*, likewise contrary, as the cause of both *coming-to-be* and *passing-away*. Since these processes go on continuously, we must look for continuous motion. Only motion in a circle is continuous, and motion in an inclined circle has the necessary duality of opposing movements. Such a motion is that of the sun along the ecliptic, which, as it approaches (spring), causes coming-to-be and, as it retreats (autumn), causes decay.

And yet Aristotle frequently calls in observed facts to substantiate his speculations. For example, in “De Caelo” (Book II, Cap. XIV), after proving, by a more or less abstract argument, that the earth is spherical, he says:

The evidence of all the senses further corroborates this. How else would eclipses of the moon show segments as we see them? . . . since it is the interposition of the earth that makes the eclipse, the form of this line (*i.e.*, the earth's shadow on the moon) will be caused by the form of the earth's surface, which is therefore spherical.

He also points to the apparent change in altitude of the stars as one travels north or south and concludes that “not only is the earth circular, but it is a circle of no great size.”

Indeed, in theory if not in his own practice, Aristotle places great emphasis on the importance of facts in connection with scientific development. In a paragraph in “De Generatione et Corruptione” (Book I, Cap. II), he says:

Lack of experience diminishes our power of taking a comprehensive view of the admitted facts. Hence, those who dwell in intimate association with

nature and its phenomena grow more and more able to formulate, as the foundation of their theories, principles such as to admit of a wide and coherent development; while those whom devotion to abstract discussions has rendered unobservant of the facts are too ready to dogmatize on the basis of a few observations.

This is surely good doctrine even for twentieth-century scientists!

An attempt to summarize Aristotle's views on physics is beyond the scope of this book, but reference may be made to two of his doctrines because of their bearing upon subsequent history.

The first is his supposed views on falling bodies. The statement is commonly made that Aristotle held that a heavy body would fall from a given height with greater velocity than a light body. It is difficult, however, to be sure from Aristotle's extant writings just what he actually held in regard to this point. The passages that seem to refer to it occur in the course of his arguments against the possibility of the existence of a void. For example, he states¹:

. . . We see the same weight or body moving faster than another for two reasons, either because there is a difference in what it moves through, as between water, air, and earth, or because, other things being equal, the moving body differs from the other owing to excess of weight or lightness. . . . And always, by so much as the medium is more incorporeal and less resistant and more easily divided, the faster will be the movement.

Here, as elsewhere, Aristotle speaks always of movement of a body through a medium. It has been suggested by Greenhill² that he may have meant *terminal* velocity, such as the constant velocity of rain drops as they approach the ground. The *heavier* drops *do* fall faster. It seems perhaps more probable that Aristotle, believing that a medium of some sort must always be present, was unaware of such distinctions as that between terminal and nonterminal velocity and actually did believe that in all stages of its motion the heavier body falls faster.³

The second doctrine referred to is that of the motion of the earth, sun, and planets. In his "De Caelo" (Book II, Cap. XIV), after a series of abstract arguments, in the course of which he states that "heavy bodies forcibly thrown quite straight upward return to the point from which they started even if they be thrown to an infinite (!) distance," Aristotle concludes "that the earth does not move and does not lie elsewhere than at the center." He supposed that the sun, planets, and stars are carried by a series of concentric spheres which

¹ This and the following quotation from Aristotle are taken from "The Works of Aristotle translated into English," vol. II, the Clarendon Press, Oxford, 1930.

² GREENHILL, *Nature*, vol. 92, p. 584.

³ Cf. COOPER, L., "Aristotle, Galileo, and the Tower of Pisa," 1935.

revolve around the earth as a center. The authority of Aristotle was so great as to render sterile the brilliant work of Aristarchus in the next century.

7. **Aristarchus** (about 310–230 B.C.) enunciated a cosmogony identical with that proposed by Copernicus 2,000 years later. No mention of this hypothesis is made in his only extant work, "On the Sizes and Distances of the Sun and Moon," but Archimedes tells us, in a book called "The Sand-reckoner," that "Aristarchus of Samos brought out a book" containing the hypothesis "that the fixed stars and the sun remained unmoved; that the earth revolves around the sun in the circumference of a circle, the sun lying in the middle of the orbit"; and that "the sphere of the fixed stars" is very great compared with the circle in which the earth revolves. The prestige of Aristotle was too great, however, and the geocentric hypothesis that he supported was so completely satisfactory to the ancient mind that Aristarchus's theory was practically lost for nearly 2,000 years.¹

8. **Archimedes** (287–212 B.C.), whose name is known to every student of elementary physics because of the famous principle of hydrostatics that bears his name, was one of the most noted physicists of antiquity. He was a man of great ability in what would now be called "theoretical (or mathematical) physics" as well as a practical engineer—a sort of ancient Lord Kelvin. In one of his books already mentioned, "The Sand-reckoner," he computes that 10^{63} grains of sand would fill the sphere of the universe as fixed by Aristarchus. In another, "On Floating Bodies," he lays the foundations of hydrostatics. His Proposition 7, in this book, enunciates the famous principle: "A solid heavier than a fluid will, if placed in it, descend to the bottom of the fluid, and the solid will, when weighed in the fluid, be lighter than its true weight by the weight of the fluid displaced."

9. **From the Greeks to Copernicus.**—To give but a passing comment to the 17 centuries between Archimedes and Copernicus would seem to give the reader the impression that no developments of moment occurred during that long period. This impression is almost correct. During ancient times Ptolemy of Alexandria (70–147 A.D.) collected the optical knowledge of his time in a book in which he discusses, among other things, reflection from mirrors—plane, convex, concave—and, particularly, refraction, which Ptolemy evidently studied experimentally. He gives, in degrees, relative values of angles of incidence

¹ Heath's "Aristarchus of Samos" is an exceedingly interesting and valuable book. It contains a review of early Greek astronomy up to the time of Aristarchus, a discussion of his work, and a translation of his only extant book, "On the Sizes and Distances of the Sun and Moon."

and of refraction for air-water, air-glass, and water-glass surfaces and describes an apparatus by which he determined these quantities; he states that for a given interface these two angles are proportional. He also mentions atmospheric refraction as affecting the apparent position of stars. He invented a complicated theory of the motions of the planets in their orbits about the earth in order to explain their apparent motions among the stars.

From Ptolemy to the Arabian Alhazen, is a span of nine centuries—twice the total lapse of time from the discovery of America to the present—during which there was stagnation in almost all lines of intellectual pursuits. But about the eighth century A.D., as an indirect result of religious activity, the Arabs began to cultivate chemistry, mathematics, and astronomy, in large part by translating into Arabic the works of the Greeks but also, in a few instances, by making original contributions. About the year 1000, Alhazen produced a work on optics in seven books. This treatise sets forth a clear description of the optical system of the eye, discusses the laws of concave and convex mirrors, both spherical and cylindrical, and carries the subject of refraction a little further by recognizing that the proportionality between the angles of incidence and refraction holds only for small angles.

During the next 500 years, a very few advances in physics were made. Roger Bacon (1214–1294), British philosopher and scientist and a monk of the Franciscan Order, taught that in order to learn the secrets of nature *we must first observe*. He believed in mathematics and in deductive science, but he clearly realized that only as these were based on observed phenomena and tested by experiment could useful knowledge result.

About the same time Petrus Peregrinus recognized that magnetic poles are of two kinds, like poles repelling and unlike attracting each other.

Then there was Leonardo da Vinci (1452–1519), Italian painter, architect, sculptor, engineer, and philosopher, whose greatness as a scientist has come to be appreciated only in recent years, for his works were left in manuscript form and were probably not widely known among his contemporaries—for which reason his influence on early science is comparatively insignificant. His belief in the value of experiment is worthy of the twentieth century: “Before making this case a general rule, test it by experiment two or three times and see if the experiment produces the same effect.” Although expressed in the vague language of his time, some of his ideas concerning what we now refer to as “force,” “inertia,” “acceleration,”

the "laws of motion," etc., were qualitatively correct. Concerning perpetual motion, he wrote: "Oh, speculators on perpetual motion, how many vain projects of the like character you have created! Go and be the companions of the searchers after gold." Rejecting the Ptolemaic theory, he held that "the sun does not move." That he was not persecuted or even burned at the stake, as was Bruno a century later, for holding such revolutionary and, therefore (!), heretical views is probably due to the fact that his doctrines were given so little publicity; for, holding no academic position, he did not teach, and he published nothing.

Finally, in the sixteenth century, the full force of that period of intense intellectual activity known as the Renaissance began to be felt in the field of physics. Then were produced such men as Copernicus, Tycho, Kepler, Galileo, Newton, who with their contemporaries and colleagues, in a space of hardly more than a century, completely broke the "spell" of Aristotle and made possible the beginnings of modern experimental science. In so far as the heliocentric theory completely revolutionized man's conception of the universe and his place in it, it is quite correct to regard the work of Copernicus as beginning a new era in scientific thought. But had it not been for other discoveries coming immediately after Copernicus, such as the telescope, Kepler's laws, Galileo's famous experiments on falling bodies, and many others, it is quite possible that the theory of Copernicus would have had the same fate as that of Aristarchus centuries earlier. It is, therefore, fitting to regard the birth of the Copernican theory as *closing* the first period in the history of physics.

10. The Copernican System.—Copernicus (1473-1543), a younger contemporary of Columbus, spent most of his life as one of the leading canons in the monastery at Frauenburg, near the mouth of the Vistula. As a student, however, he had studied, among other subjects, mathematics and astronomy; at the University of Bologna, he was a pupil of a famous astronomer, Novara, who supported the Pythagorean system of the universe as against the Ptolemaic. Copernicus's new theory of the universe is set forth in his "*De Revolutionibus Orbium Coelestium*," which he allowed to be printed near the close of his life. The new theory appears to have been the result of long reflection upon the difficulties of the current astronomical theory of his day, and upon the various speculations of ancient philosophers.

Copernicus perceived that, by assuming that the earth is a planet like the others and that all the planets move in circles around the sun, a great simplification, both philosophical and mathematical.

could be made with regard to the world system. He could thus easily account for the seasons and for the apparent retrograde motion, at times, of the planets. The rotation of the earth on its axis causes the *apparent* daily motion of the sun, moon, and stars; and he pointed out that, probably, the stars are too far away for any motion of the earth to affect their apparent positions. He gave the correct order of the planets from the sun outward.

Whatever the system as proposed by Copernicus lacked quantitatively, it was correct, in its main outline, qualitatively. Its reasonableness set a few men thinking and did much to usher in a new era in science, an era that could come only when truth could have the opportunity of standing alone, *unaided* or *unhindered* by the "authority" of 2,000 years.

SECOND PERIOD (1550-1800 A.D.): RISE OF THE EXPERIMENTAL METHOD

11. **Galileo Galilei** (1564-1642).—Galileo is widely, and quite correctly, regarded as the father of modern physics. To be sure, modern physics has grandfathers and still more remote ancestors, but none of them gave to experimental physics so much as Galileo. Physicists should be acquainted with the main details not only of his scientific work but also of his life. The student is urged, therefore, to peruse one of the several biographies and to read some of Galileo's writings, at least in translation. Even a short time spent in following his deductions and in reading *firsthand* something he wrote will prove both interesting and instructive.

Galileo was descended from a noble family, and it is quite probable that he inherited from his father the spirit of free inquiry which characterized his life. For, in the writings of the elder Galileo, who was well educated and was an accomplished musician, is the statement: "It appears to me that they who in proof of any assertion rely simply on the weight of authority, without adducing any argument in support of it, act very absurdly."

As a student in the monastery of Vallombrosa, near Florence, the young Galileo excelled in the classics and literature and was something of a poet, musician, and art critic. He also showed an aptitude for science and exhibited considerable mechanical inventiveness. At the age of seventeen he was sent to the University of Pisa to study medicine. It was here that he made his first discovery and invention. One day, in 1581, he noticed the regular oscillations of the great hanging lamp in the cathedral at Pisa. Although the amplitude of these oscillations became less and less, they were all

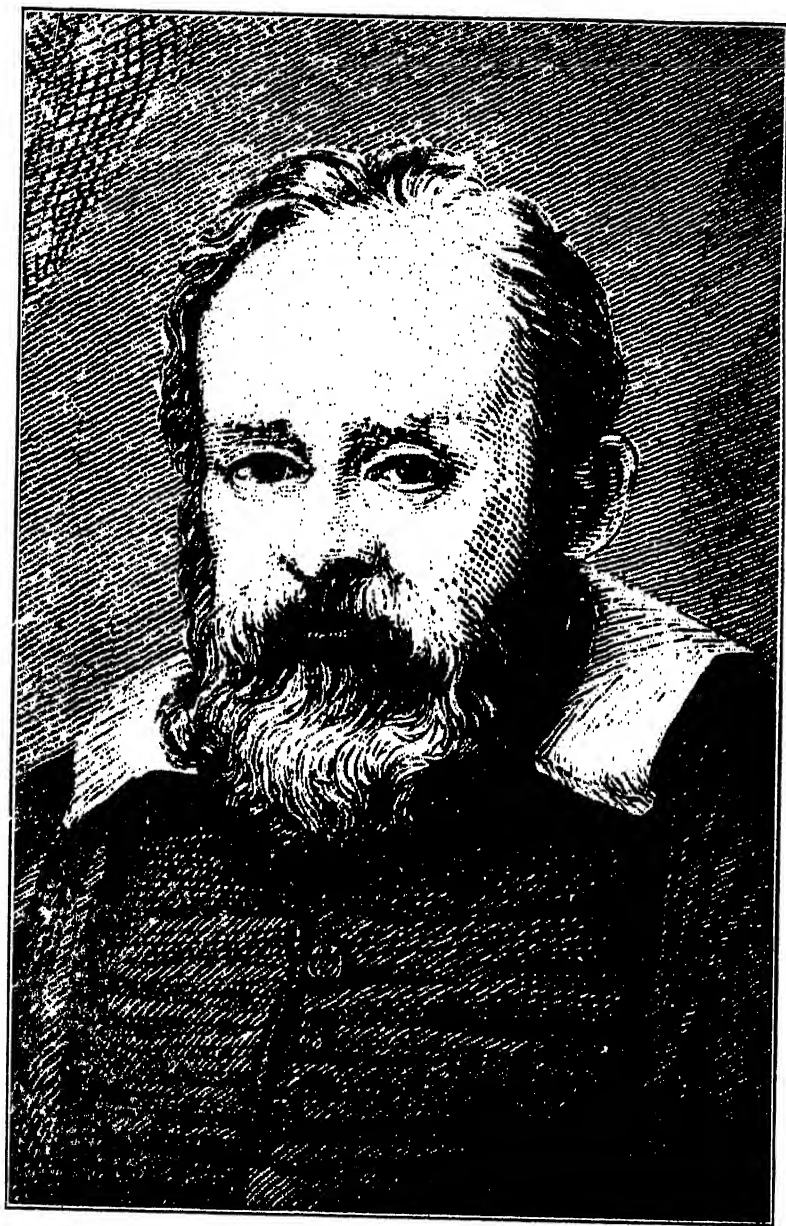


PLATE 1.—Galileo.

performed in the same time, as he determined by counting his pulse. Turning the process around, he invented a "pulsometer," a ball-and-string (*i.e.*, simple pendulum) device, whose length, when adjusted to synchronism with the pulse, was a measure of its frequency.

But the urge toward mathematics and science overcame the pecuniary advantages of a medical career. At the age of twenty-six, Galileo became Professor of Mathematics at Pisa. Here he began a systematic investigation of the mechanical doctrines of Aristotle. He soon demonstrated by experiment to his own satisfaction that Aristotle was in error in many of his assertions, and these errors he proclaimed energetically from his professorial chair.

It will be recalled that Aristotle was commonly understood to teach that a heavy body falls faster than a light one. This doctrine had been questioned, on the basis of actual test, by various writers, *e.g.*, by Philoponus in the fifth century and by Benedetto Varchi in the generation before Galileo. Nevertheless the authority of Aristotle had continued to be accepted. To test the point, Galileo apparently tried the famous experiment of dropping bodies of unequal weight from the leaning tower of Pisa and found that they all fell with practically equal velocities. We know nothing of the details of the experiment; indeed it is not even certain that he performed it at all.¹ It is certain, however, that Galileo and a few of his friends were convinced of Aristotle's error; and it seems equally certain that the majority maintained, in spite of all experiments and arguments to the contrary, that Aristotle must be right.

And then began a persecution which was to last Galileo's lifetime, increasing in severity as he grew older and, finally, resulting in imprisonment. To present the details of his stormy life is far beyond the scope of this book. The reader is referred to his biographers.²

He was soon forced to quit Pisa, and, in 1592, he became Professor of Mathematics at the University of Padua, where he remained 18 years, enjoying comparative liberty of thought and teaching. His fame as a teacher spread all over Europe, and his lectures were crowded.

In 1608, a Dutch optician, Lipperhey, as a result of a chance observation of an apprentice, had succeeded in "making distant objects appear nearer but inverted" by looking through a tube in which were mounted two spectacle lenses. News of this invention

¹ Recently there has been a tendency to question the historical accuracy of the various accounts of this experiment. For an excellent discussion of the whole subject see L. Cooper, "Aristotle, Galileo and the Tower of Pisa," 1935.

² FAHIE, J. J., "Galileo: His Life and Work," 1903; BREWSTER, SIR DAVID, "Martyrs of Science," 1870; LODGE, OLIVER, "Pioneers of Science."

reached Galileo in June, 1609. Immediately grasping the principle involved, he made a telescope and exhibited it in Venice "for more than a month, to the astonishment of the chiefs of the republic." By January, 1610, Galileo had made a telescope with a power of 30 diameters,¹ and with this instrument he made a number of fundamental discoveries. He saw that the number of fixed stars was vastly greater than could be seen by the unaided eye, and thus he was able to explain the agelong puzzle of the Milky Way. He saw that the planets appeared in his telescope as luminous disks while the stars still remained points of light, and he discovered the satellites of Jupiter. These discoveries naturally made Galileo famous, and he soon accepted an invitation to return to Pisa as "First Mathematician and Philosopher," at a very substantial increase in salary, but at a sacrifice, unfortunately, of his "academic freedom" in Padua. Continuing his astronomical investigations, he discovered the crescent phases of Venus, sunspots and the rotation of the sun, the faculæ of the solar atmosphere, and the libration of the moon. In 1612, he published his "Discourse on Floating Bodies."

At first, it seemed as if his fame had silenced all opposition from the church. But the support that his discoveries gave to the hated Copernican theory and his vigorous attacks on Aristotelian philosophy roused his enemies to fury, with the result that in 1615 he was hauled before the Pope and, under threat of imprisonment and torture, was "enjoined . . . to relinquish altogether the said opinion that the sun is the center of the world and immovable . . . nor henceforth to hold, teach, or defend it in any way . . . " Simultaneously, it was decreed that the works of Copernicus "be suspended until they be corrected." Galileo acquiesced in these decrees and was allowed to return to Pisa, where he continued his researches along such lines as would not give offense.

In 1623, one of Galileo's friends, Barberini, became Pope Urban VIII, from whom Galileo received assurances of "pontifical good will." Thereupon, thinking that the decree of 1615 would no longer be enforced, he began the writing of his great book, "Dialogues on the Ptolemaic and Copernican Systems." This was published in 1632, under formal permission from the censor. The form of these dialogues is ingeniously contrived to abide by the *letter* of the decree of 1615. Three "characters" carry on the discussion: Salviati, a Copernican; Simplicio, an Aristotelian; and Sagredo, a witty, impartial, good-

¹ Galileo's telescopes were similar to the modern opera glass—a double-convex (or plano-convex) object glass and a double-concave eyepiece. Thus, they had an erect image.

natured chairman. The dialogues cover four "Days," during which the arguments for and against each system are set forth with apparent impartiality and without reaching any *stated* conclusion. Nevertheless, the general effect of the book was "a powerful plea for Copernicanism."¹

In spite of its enthusiastic reception by the public, the form of the book did not deceive his enemies, who were now determined that he must be silenced. This was effectually accomplished by representing to the Pope that the Simplicio of the dialogues, whose ignorance was very apparent, was simply a caricature of the Pope himself. In spite of the absurdity of the argument—for Galileo would hardly have risked offending Urban VIII, his one friend in the church—the Pope, who seems to have been both ambitious and vain, became convinced that Galileo "had made game of him." Whereupon he was ready to join Galileo's enemies in persecuting that great scientist, ostensibly for "the safety of the church and the vindication of its decrees."

This tragic incident is illustrative of the fact that, in the seventeenth just as in the twentieth century, much of the "warfare between science and theology"—as Andrew D. White calls it²—has been based upon personal motives rather than upon a sincere wish to uphold theological doctrines. Intolerance, even when it is sincere, is to be condemned; but *insincere* intolerance is to be despised. Yet, one cannot fully comprehend the forward march of science unless one recognizes the seriousness of the obstacles of this kind that have had to be overcome. Great as is the fame of Galileo, how much more might he have accomplished if the energy that he was forced to spend in overcoming opposition could have been directed toward his researches. Without doubt, the giant strides being taken by science in the twentieth century are possible only because of the freedom that the scientist now enjoys.

Galileo was presently called before the Inquisition. He was sixty-seven years old, impaired in health and in spirit. Bowing to the inevitable because of the magnitude of the forces arrayed against him, he followed the advice of his friends and indicated his "free and unbiased" willingness to recant, to "abjure, curse, and detest the said heresies and errors and every other error and sect contrary to the Holy Church," and he further agreed "never more in future to say or assert anything, verbally or in writing, which may give rise to a similar suspicion." Thereafter Galileo was kept a prisoner under suspended

¹ FAHIE, *loc. cit.*

² WHITE, ANDREW D., "The History of the Warfare between Science and Theology."

sentence, first at Rome, then at his home in Arcetri. Here, during the last years of his life, he prepared and, in 1636, published his "Dialogues on Two New Sciences"¹ (*i.e.* Cohesion and Motion).

The dialogues on "Motion" sum up Galileo's earlier experiments and his more mature deliberations. He states that "if the resistance of the media be taken away, all matter would descend with equal velocity." He deduces the formulas of uniformly accelerated motion. He shows that the path of a projectile is parabolic under suitable limiting conditions and states that, if all resistance were removed, a body projected along a horizontal plane would continue to move forever. His work on mechanics paved the way for the enunciation by Newton of the famous three laws of motion, which form the foundation of mechanics.

12. Tycho Brahe (1546–1601) and Kepler (1571–1630).—The work of Tycho and Kepler is particularly interesting not only because of its direct bearing on the development of physics but more particularly because of the mutual dependence of the work of each one upon that of the other, a relation very common in present-day science. Tycho was the experimentalist, the observer, who supplied the accurate data upon which Kepler, the theorist, built a new theory of planetary motion. Without a Kepler to build a theory from them, Tycho's observations would have attracted hardly more than passing notice. Kepler, in turn, might have theorized to his heart's content, but, without the accurate data of a Tycho, those theories would ultimately have shared the fate of Aristotle's. Sometimes theory precedes, sometimes experiment. But neither can get far without the other.

Tycho Brahe, born of a noble family in Sweden, was educated for a career as a statesman, but he developed a consuming interest in astronomy. By means of observations of his own, he soon found that the current astronomical tables were incorrect. In 1575, he was put in charge of the observatory of Uraniborg by King Frederick II of Denmark, one of his duties being to make *astrological* calculations for the royal family. Here he spent 20 years making systematic observations of the planets, constructing a star catalogue, and accumulating

¹ See the excellent translation by Crew and de Salvio.

Some writers have severely censured Galileo for yielding to the Inquisition. They say that "had Galileo added the courage of a martyr to the wisdom of a sage . . . science would have achieved a memorable triumph" (see Brewster, "Life of Newton"). Whatever opinions on this question one may hold, one fact stands out indisputable: Had Galileo *not* yielded he would surely have been cast into the dungeon and would probably have been burned at the stake. We should not then have had handed down to us these dialogues on Motion, so fundamental to our modern physics.

other astronomical data, always with the highest possible order of accuracy that could be attained without a telescope. In 1599, Tycho undertook to establish a new observatory at Prague for the German emperor, Rudolph the Second, but in the midst of this work he suddenly died (1601).

Now among Tycho's assistants at Prague had been, during the last few months, a brilliant young mathematician, Johann Kepler. He succeeded Tycho as principal mathematician to the emperor and undertook to carry to completion the new set of astronomical tables based on elaborate observations, which Tycho had begun. Kepler remained at Prague until 1612; from then until his death in 1630, he held a professorship at Linz.

Kepler believed thoroughly in the Copernican system, which Tycho had rejected for a geocentric system of his own. It is one of the dramatic situations in science that Tycho's data on planetary motions, taken in support of his own theory, became, in the hands of Kepler, the clinching argument for the Copernican system. Using Tycho's observations, Kepler made a special study of the motion of Mars. He tried to reconcile the various recorded positions of the planet by assuming circular orbits for Mars and for the earth, trying various positions of these orbits relative to the sun. None worked. By resorting to the Ptolemaic notion of epicycles and deferents, some improvement resulted, but still the observed positions differed from the computed, in some cases by as much as 8 minutes of arc. Kepler knew that Tycho's observations could not be in error by that amount. Some new concept regarding planetary motion was necessary.

Then Kepler gave up *uniform* circular motion and assumed that the speed varied inversely as the planet's distance from the sun. This assumption is his famous "second law," that *the radius vector from the sun to the planet describes equal areas in equal times*. It worked approximately, but still there were systematic errors which exceeded the possible errors of observation. Finally, he cast aside the last traditions of the Ptolemaic system and tried orbits of other forms, first, an oval path, and, then, an ellipse, with the sun at one focus. At last, his years of computation bore fruit. The path *was* an ellipse. Theory and observation agreed! And one of the most important and far-reaching laws in all science had been discovered, all because of a discrepancy of 8 minutes of arc between observation and theory! In fact, one of the striking things in the growth of science, particularly physical science, is the fact that many fundamental advances have come about because of just such discrepancies, frequently very small ones, between observation and theory. Studying Tycho's observa-

tions further, Kepler finally hit upon the true relation between the periods of the planets and the radii of their orbits, a relation now known as Kepler's "third law," that *the squares of the times of revolution around the sun are as the cubes of the mean orbital radii*.

Thus were completed the three laws of planetary motion that Kepler handed down to posterity and which, sweeping away all remnants of the Ptolemaic system, paved the way for modern astronomy:

1. The planets move around the sun in orbits which are ellipses, with the sun at one focus.
2. The radius vector (from sun to planet) sweeps over equal areas in equal times.
3. The squares of the periods of revolution of the planets around the sun are proportional to the cubes of the (mean) radii of their respective orbits.

But what makes the planets move? *Why* do the outer ones go more slowly? Is there "one moving intelligence in the sun, the common center, forcing them all around, but those most violently which are nearest?" Kepler speculated long on this question and arrived at the idea of an attraction acting between any two material bodies. This *qualitative* idea of Kepler's was later developed by Newton into his *quantitative* theory of universal gravitation. Kepler himself, however, seems to have had no idea that it is this very attractive force which keeps the planets themselves in their orbits.

In passing, it may be mentioned that Kepler also made substantial contributions in the field of optics. He understood clearly the principle of total reflection and how to determine what we now call the "critical angle." He studied atmospheric refraction as affecting the apparent position of the heavenly bodies and worked out an approximate formula to allow for this error for all positions, from zenith to horizon. He was the first to propose the meniscus type of lens. And he proposed the Keplerian or astronomical type of telescope, in which a *real* image is formed, thus making possible accurate measurements by means of cross hairs in the focal plane of the objective.

13. The Experimental Method Spreads. The impetus given to science by that great trio, Galileo, Tycho, and Kepler, resulted in an ever increasing number of investigators in the generations that followed. We can mention only a few of them. Of great significance, too, is the fact that at about this time there were formed in several European centers various learned societies which brought together, for argument and discussion, men of kindred interests. The Lincean Society was founded in Italy, in 1603; the Royal Academy of Sciences,

in France, in 1666; and the Royal Society for the Advancement of Learning, in England, in 1662. The continued improvement of the art of printing enormously facilitated the diffusion of scientific knowledge.

In 1600, Gilbert, an English physician, published his famous work, "De Magnete," based largely upon his own experiments, in which he showed the fallacy of such popular fancies as the belief that lodestones lost their magnetic power when rubbed with garlic and regained it again when smeared with goat's blood. He was the first to recognize that the earth is a great spherical magnet, and he actually magnetized a small sphere of iron and showed that it behaved as the earth behaves.

Among other workers in magnetism may be mentioned Kircher (1601-1680), who, by measuring the force required to pull a piece of iron from either pole of a magnet, demonstrated the equality of the two poles; Cabeo (1585-1650), who showed that an *unmagnetized* iron needle floated freely on water would place itself along the earth's magnetic meridian; and Gellibrand (1597-1637), who discovered the secular variation of the magnetic declination.

In the field of optics, there was Scheiner (1575-1650), who studied the optics of the eye; Snell (1591-1626), who discovered the true law of refraction; and Cavalieri (1598-1647), who gave the correct formula for the focal length of a thin glass lens in terms of the radii of curvature of the two sides. Studies in acoustics, also, were not wanting. For example, Mersenne (1588-1648), after having investigated the laws of vibrating strings, determined, in absolute measure, the frequency of a tone. He also measured the velocity of sound by observing the time interval between the flash of a gun and the arrival of the report.

In the field of fluid mechanics, there was Torricelli (1608-1647), who studied the flow of liquids from orifices, discovered the principle of the barometer, and observed variation in barometric height with altitude. Working independently, Guericke (1602-1686) invented the air pump. Pascal (1623-1662) measured the difference in barometric height between the base and the top of a mountain, correctly explaining the reason for the difference, and, later, announced the famous principle of hydrostatics that bears his name.

Not only was physics, as a subject, beginning to assume definite form, but even the usual subdivisions such as mechanics, light, sound, etc., were beginning to crystallize out.

Then came a man

... towering head and shoulders above all his contemporaries, a veritable giant among the giants; a man whose intellect and whose contributions to

knowledge are incomparably greater than those of any other scientist of the past, that prince of philosophers, Sir Isaac Newton.¹

The other "giants" referred to, contemporaries of Newton, are such men as Boyle, Huygens, and Hooke.

14. Sir Isaac Newton (1642-1727).—Newton was born in the little hamlet of Woolsthorpe, England, in 1642, less than a year after the death of Galileo. In the public school at Grantham, he showed at first no exceptional aptitude for his studies, but ultimately he rose to the highest place in the class. Then, at the age of fifteen, he was removed from school to assist his widowed mother in running the family estate at Woolsthorpe. But he had little taste for farming. Rather, he was interested in studying and in devising various mechanisms. He made a water clock, water wheels, sundials, and a working model of a windmill. And one morning his uncle found him under a hedge studying mathematics when he should have been farming. Thereupon Newton's mother wisely decided that an educational career was more suitable for her son, and he was sent back to school and, ultimately, to Cambridge, which he entered in 1661. Here his fondness for mathematics continued, and soon his creative genius began to appear. While still a student, he discovered the binomial theorem, developed the methods of infinite series, and discovered "fluxions," or the differential calculus.

Soon thereafter, an outbreak of the plague closed the University for some months, during which time Newton, at the family estate at Woolsthorpe, began his speculations on the subject of gravity, which later led to his enunciation of the inverse-square law. It was here that the much-quoted "falling apple" episode is said to have occurred, which is supposed to have given Newton the basic idea of *universal* gravitation. But Newton himself makes no mention of the incident, and it seems far more probable that at Cambridge he had read Kepler's qualitative proposal of a general principle of gravitation. Certainly, Newton was familiar with the three laws of planetary motion that Kepler had discovered.

In 1667, Newton returned to Cambridge as Fellow of Trinity. Two years later, at the age of twenty-six, he was appointed Lucasian Professor of Mathematics, a chair which he held for nearly 30 years. In 1703, he resigned his professorship to devote himself to his duties as Master of the Mint, to the scientific work of his contemporaries, and to defending his own work against the attacks of jealous rivals. In this same year, he was elected President of the Royal Society, an

¹ HART, "Makers of Science."

office which he held until his death. In 1705, he was knighted by Queen Anne.

Most of Newton's important scientific work was done before he vacated the professorship, although he remained thereafter "a power of the first magnitude in the world of science." In his later years, he devoted much time to theological studies. He died Mar. 20, 1727, at the ripe old age of eighty-five. Throughout his life, he shunned publicity and retained a modesty and simplicity which are indicated by a sentiment uttered shortly before his death:

I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.

Any brief account of Newton's work must inevitably give a very inadequate impression of his contributions to science. The student is urged to read some of his writings firsthand or, at least, some extensive biographical discussion of his life and work. We can here refer only to a very few of his researches on optics and on mechanics.

Newton's work on *optics* arose out of an attempt to improve lenses. The inability of a lens with spherical surfaces to bring parallel rays to a point focus was early recognized. In 1629, Descartes had shown that lenses with hyperbolic or, under certain conditions, parabolic surfaces should be free from the defect which we now call "spherical aberration." Newton found, however, that such lenses produced only a very slight improvement in the image, and he conjectured that, perhaps, the trouble lay not in the lens but in the light itself.

Accordingly, he procured a prism of glass and, placing it over a hole $\frac{1}{4}$ inch in diameter through which sunlight was shining into a darkened room, he observed the "very vivid and intense colors" produced on a wall some 20 feet distant. Newton was surprised to find that this "spectrum," as we now call it, was so much longer than it was wide ($13\frac{1}{4}$ by $2\frac{5}{8}$ inches). The *width* subtended at the hole an angle corresponding *exactly* to the sun's angular diameter, but the length could not be so explained. He made various surmises as to the origin of the colors, such as the varying thickness of the prism from apex to base, the unevenness of the glass, a curvilinear motion of the light after leaving the prism, etc. One by one, experiment proved these hypotheses wrong. Finally, he isolated one ray, or "color," by suitable screens and caused it to pass through a second prism. In this way, he could measure the refrangibility of a single ray. And he found that the refrangibility increased from red to violet, that,



PLATE 2.—Newton.

therefore, the first prism simply "sorted out" the colors, which, in combination, made "white" light. In other words, he discovered that so-called "white light" is made up of the spectral colors—a very elementary concept to us of the twentieth century but very new and of far-reaching importance in 1666.

Newton at once saw that this dispersion of light was the cause of his failure to effect any substantial improvement in telescopes by use of paraboloidal lenses. Furthermore, he concluded, on the basis of a hurried experiment, that, in different mediums, dispersion was always proportional to refracting power. If this were so, it would follow that by no combination of lenses of different materials could chromatic aberration be eliminated. This singular error of Newton's retarded the development of refracting telescopes for many years. In 1730, Hall made several achromatic combinations of crown and flint glasses, but he published no account of his work, so that, when Dolland, about 1757, rediscovered the method of making such combinations, he was able to secure a patent on it—an invention that had been within the grasp of Newton three-quarters of a century before.

Newton's theories concerning the nature of light are of great historical interest. And much has been written concerning the extent to which he is supposed by some to have retarded the development of optics, by espousing the corpuscular theory as against the wave theory of his contemporaries, Huygens (1629–1695) and Hooke (1635–1703). Accordingly, it may be of interest to point out, by a quotation or two, that Newton was by no means dogmatic in his support of the corpuscular theory and that later writers may have taken him, in this regard, more seriously than he intended.

In a communication to the Royal Society, in 1675, concerning "An Hypothesis Explaining the Properties of Light," Newton states:

. . . I have here thought fit to send you a description . . . of this hypothesis . . . though I shall not assume either this or any other hypothesis, not thinking it necessary to concern myself whether the properties of light discovered by men be explained by this or any other hypothesis capable of explaining them; yet while I am describing this, I shall sometimes, to avoid circumlocution . . . speak of it as if I assumed it.

He then proceeds to describe "an ætherial medium, much of the same constitution with air but far rarer, subtler and more strongly elastic" and supposes that

. . . light is neither æther, nor its vibrating motion, but something of a different kind propagated from lucid bodies. They that will may suppose it an aggregate of various peripatetic qualities. Others may suppose it multi-

tudes of unimaginable small and swift corpuseles of various sizes springing from shining bodies . . . and continually urged forward by a principle of motion which in the beginning accelerates them, till the resistance of the ætherial medium equals the force of that principle much after the manner that bodies let fall in water are accelerated till the resistance of the water equals the force of gravity.

Nearly 33 years later (1704); Newton published his optical researches in book form in his well-known "*Opticks*," bringing out a third edition in 1721. His first sentence reads: "My Design in this Book is not to explain the Properties of Light by Hypotheses, but to propose and prove them by Reason and Experiment." He then, in some 300 pages, gives his researches on refraction, reflection, colors of thin plates, etc., and he concludes the book by "proposing only some queries in order to further search to be made by others." One of these queries expresses his objections to the wave theory of light:

28. Are not all hypotheses erroneous in which light is supposed to consist in pression or motion propagated through a fluid medium? If light consists only in pression propagated without actual motion, it would not be able to agitate and heat the bodies which refract and reflect it, and . . . it would bend into the shadow. For pression or motion cannot be propagated in a fluid in right lines beyond an obstacle . . . but will bend and spread every way into the quiescent medium which lies beyond the obstacle.

In order to account for the colors of thin films, which are now regarded as strong evidence for the wave nature of light, he supposes that "every ray of light in its passage through any refracting surface is put into a certain transient constitution or state, which in the progress of the ray returns at equal intervals and disposes the ray at every return to be easily refracted through the next refracting surface and between the returns to be easily reflected by it." He even suggests that the effect might be due to vibrations excited by the "rays" in the material medium which vibrations "move faster than the rays so as to overtake them; and that when any ray is in that part of the vibration which conspires with its motion, it easily breaks through a refracting surface, but when it is in the contrary part of the vibration which impedes its motion, it is easily reflected . . . But whether this hypothesis be true or false, I do not here consider."¹

It is clear that Newton regarded his corpuscular theory as a tentative one, subject to confirmation or rejection on the basis of further experiments. If his theory really did retard progress in optics, the fault lay rather with those who attached too great weight to his opin-

¹ Quoted in Preston's "*Light*," p. 22.

ions, perhaps because they retained something of that medieval respect for authority which had kept Aristotle enthroned so long.

Newton's speculations may also serve as an example to illustrate the rule that even the greatest intellect works on the basis of the knowledge and viewpoints of its age. Had Newton lived a century later, he would probably have been one of the first believers in the wave theory. The fact that great scientists share the limitations of their age is an added reason for treating their speculative opinions chiefly as sources of inspiration for further experiment.

Newton's researches on optics alone would have given him a high rank among the scientists of his time. But of still greater value was his work in *mechanics*. In announcing that "every particle of matter in the universe attracts every other particle with a force inversely proportional to the square of the distance between the two particles," in showing that this one universal and comparatively simple law governs not only the motions of the planets around the sun and of the satellites round their planets but, probably, also the relative motions of all the heavenly bodies, he gave to the world a truth which exercised an enormous influence upon thought. This achievement of Newton's played a large part in bringing about the general conviction that the physical universe in its entirety is governed by law, not by caprice.

Newton himself has told us how he came to discover the law of gravitation. First he attacked the problem of finding a law of attraction between two bodies, such as the sun and a planet, which would result in Kepler's third law, *viz.*, that squares of the periods of rotation of the planets around the sun are proportional to the cubes of their mean distances from the sun. He found that a gravitational attraction varying as the inverse square of the distance gives this law of planetary motion.

Then he saw that a test of this inverse-square law could easily be made by comparing the acceleration of the moon toward the earth with the acceleration of falling bodies at the surface of the earth. It was known that the distance between the moon and the earth's center is about sixty times the earth's radius. By the inverse-square law, therefore, the moon should "drop" toward the earth, in 1 second, $\frac{1}{60}^2$ as far as a body at the surface of the earth drops in 1 second. The latter distance being, from observations on falling bodies, 16 feet, the former should be $\frac{16}{60^2}$ feet, or 16 feet in 1 minute. But the acceleration of the moon could be determined directly by applying to the moon the expression he had used for the motion of the planets in their orbits, *viz.*,

$$a = \frac{v^2}{r} = 4\pi^2 \frac{r}{T^2},$$

where v is the velocity of the moon in its orbit or T is the period of its motion around the earth and r is the radius of the orbit. Now r is equal to sixty times the earth's radius, which was taken as 3,436 miles, on the then common assumption that a degree of latitude is 60 miles. On this basis, the moon is found to "drop" 13.9 feet toward the earth in a minute, instead of 16 feet, as should be the case if the inverse-square law were obeyed. Thus the two results did not agree.

Newton was twenty-three years old at the time, and he laid this calculation aside, not mentioning it to anyone. Some years later, however, he learned of a new and more accurate determination of the length of a degree which had been made by Picard, who found not 60 but more nearly 70 miles. In the meantime Newton had also succeeded in proving that a homogeneous sphere attracts an external body as if all its mass were concentrated at its center, thereby removing one small uncertainty in the calculation. So, on the basis of Picard's new value for the length of a degree, Newton revised his computations on the moon's acceleration and, to his great joy, found that it falls toward the earth 16 feet in a minute, just as predicted by the inverse-square law. At last, he had discovered the true law of gravitation. On the basis of this law, he could now *derive* all three of Kepler's laws. Theory and observation checked perfectly.

These results, together with some propositions on the motion of the planets, were communicated in 1683 to the Royal Society, which requested permission to publish Newton's complete researches on the subject of motion and gravitation. In 1687, there appeared the first edition of his "*Principia*" or, in full, "*Philosophiæ Naturalis Principia Mathematica*" (Mathematical Principles of Natural Philosophy), "without exception the most important work in natural philosophy extant."¹ The original is in Latin, but English translations are available. The treatise is divided into three books, the subject matter of each being presented by propositions, theorems, and corollaries. The first two books deal with general theorems concerning the motions of bodies, whereas the third contains applications to the solar system. The entire treatise is characterized by the exposition of the principle of universal gravitation and its ramifications but, as the author carefully points out, without attempting any hypothesis as to the *cause* of gravitation.

In this treatise the famous three laws of motion are assumed as axioms. Their greatest merit lay in the fact that they contain just enough to constitute a complete basis for the science of mechanics and no more. The laws can be expressed with slightly greater clarity in modern terminology, and we now realize that in part they express

¹ HART, *loc. cit.*

in *March* and farthest north in *September*, the angular distance between the two positions being about 40 seconds of arc. Bradley continued his observations on other stars, and, in 1728, he came to the conclusion that the observed displacement was not due to parallax at all but to an apparent shift in the star's position due to a combination of the velocity of light with that of the earth in its orbit.¹ He was thus enabled to deduce a value for the velocity of light. The value so found was in substantial agreement with that determined by Römer a half-century earlier, from a study of the motion of Jupiter's moons, which constituted the first determination of the velocity of light. This discovery of Bradley's was the first in the series that formed the basis for the modern theory of relativity.

Theories as to the nature of light made no material progress during the eighteenth century. Some writers are inclined to ascribe this to the prestige given to the corpuscular theory by the fact that it was supported by Newton, whose preeminence "seemed to act like a spell," as had Aristotle's centuries before. But if so, then, likewise, the kinetic theory of heat held by Newton should have been uppermost during the eighteenth century, whereas, as has already been pointed out, the reigning theory of heat during this period was the caloric theory. Lack of progress in the theory of light was more likely due to the lack of any crucial experiment, just as was the case with theories of heat. Science has never progressed on the basis of speculation only.

19. Electricity during the Eighteenth Century.—Electricity received a great deal of attention during the eighteenth century, but research was concerned principally with electrostatics. Stephen Gray (1670–1736) distinguished clearly between conductors and nonconductors of electricity and showed that even conducting bodies may be electrified provided they are insulated. Du Fay (1698–1739) showed that flames exercise a discharging power and that there are two kinds of electricity, which he called "vitreous" and "resinous." He was thus led to propose the two-fluid theory of electricity. During the first half of the eighteenth century, the electroscope was invented (by Hawksbee, in 1705), frictional electric machines were developed, the Leyden jar was discovered (1745), and there was considerable popular interest in electrical phenomena. During the latter half of the century, three names stand out preeminent: Benjamin Franklin (1706–1790), Henry Cavendish (1731–1810), and Charles A. Coulomb (1736–1806).

Franklin's experiments began about 1745. One of his first observations was the effect of points "in drawing off and throwing off the electrical fire." He proposed the one-fluid theory of electricity, some-

¹ See any text on astronomy for further explanation of aberration.

what similar to the caloric theory of heat. This theory supposed that all bodies contained naturally a certain amount of the electric fluid. When a body had an excess of the fluid, it exhibited one of the two kinds of electrification, which Franklin chose to call for this reason *positive*; when it had a deficit, it exhibited the other kind, which he called *negative*. Certain features in the appearance of electric sparks led him to identify his positive electrification with that which had been called vitreous. (In the light of the modern electron theory, we should prefer that he had applied the signs in the opposite order!)

About 1750, Franklin began to speculate on the identity of electricity and lightning, pointing out many similarities and proposing, by means of a pointed iron rod, to "draw off the fire from a cloud." Franklin's writings were published in Europe, and in 1752 Dalibard tried the experiment in Paris, confirming Franklin's prediction. A short time later, Franklin performed the famous kite experiment, so well known to every schoolboy; this led to his study of atmospheric electricity and to his invention of the lightning rod. Franklin's researches occupied but a small portion of his long and busy life, but they were sufficient to give him a high standing among the scientists of the world.

Quantitative researches in electricity began with Cavendish and Coulomb. Cavendish is known not only for his work in electrostatics but also for his researches in chemistry and for the well-known "Cavendish experiment," in 1798, in which he determined the constant of gravitation. His electrical researches were very extensive, but most of his work remained unknown, for he published only one paper of importance. He left behind a large amount of material, however, in the form of manuscript notes, and these were edited and published in 1879 by Maxwell. In these experiments, Cavendish proved the inverse-square law of electrostatic force; measured capacity, recognized the principle of the condenser, and measured the specific inductive capacity of several substances; arrived at a reasonably clear idea of the quantity which we now call "potential"; and anticipated Ohm's law by 50 years. Had these important measurements been communicated to his scientific contemporaries, the history of electricity might have been substantially modified. As it is, the credit of *discovery* seems fairly to belong to others, for a discovery is of no importance to any one else if it is kept secret.

Coulomb's work in electricity grew out of his development of the torsion balance, originally used for studying the torsional elasticity of wires. In the period 1785-1789, he published seven papers on electricity and magnetism in the *Mémoires de l'Académie Royale des*

Sciences. In these papers, he showed by means of the torsion balance that electrostatic forces obey the inverse-square law; that, on conducting bodies, the charge exists only on the surface; and that the capacity of a body is independent of the nature of the material of which it is composed. He advocated the two-fluid theory of electricity.

20. Close of the Second Period.—The end of the eighteenth century found rival theories contending in each of three of the subdivisions of physics: the caloric *vs.* the kinetic theory in heat; the corpuscular *vs.* the undulatory theory in light; and the one-fluid *vs.* the two-fluid theory in electricity. The very fact that these issues were raised in rather clean-cut fashion is an indication of the tremendous strides that had been taken since Galileo. But most important of all, men had learned the value of experiment and observation and the fallacy of blindly following "authority."

During the next century the caloric theory was definitely ruled out, apparently never to return. The corpuscular theory of light seemed likewise to be effectively disproved by many new experiments. But further progress in regard to the nature of electrification waited, for the most part, until the twentieth century.

THIRD PERIOD (1800-1890 A.D.): THE RISE OF CLASSICAL PHYSICS

21. The Nineteenth Century in Physics.—So much was added to physical knowledge during the nineteenth century that an adequate history of this period would almost constitute a textbook of physics. We can only comment briefly upon the principal lines of advance and a few of the most important discoveries, selecting especially those that form the background for the characteristic advances of the present century.

In mechanics, there was Hamilton, who discovered in the "Hamiltonian equations" a new form of the equations of motion which is especially convenient in attacking theoretical questions. The theory of the motion of rigid bodies, including the gyroscope, was worked out as well as the mathematical theory of elasticity. The subject of hydrodynamics, dealing with the motion of fluid of all sorts, was developed. In dealing with the flow of viscous fluids, however, only simple problems could be solved; further study of such fluids, by half-empirical methods, was not made until during the present century, after the invention of the airplane.

The work in other fields was more striking. The most significant discoveries and advances were: the establishment of the kinetic theory of heat, and the development of the kinetic theory of gases; the

victory(?) of the wave theory of light over the corpuscular theory; the formulation of the general law of the conservation of energy; the discovery of the second law of thermodynamics; and, above all, the discovery, by Faraday and others, of the whole range of electromagnetic phenomena, culminating in Maxwell's theory of the electromagnetic field and of light.

Of these lines of advance we shall select three for detailed presentation, choosing those which bear more or less directly upon modern developments in physics; and as typical scientists of the period, we shall discuss especially Faraday and Maxwell.

22. Heat and Energy.—The law of the conservation of energy is one of the most fundamental and far-reaching of all physical laws, and yet, curiously enough, it is of comparatively recent origin, for it was not announced until the middle of the nineteenth century. As exemplified in mechanics, it had been recognized during the eighteenth century, in the theory of the "vis viva"; but its announcement as a general law of physics awaited experimental work demonstrating the definite equivalence of heat and mechanical work.

The first *qualitative* experiment bearing on the nature of heat was performed in 1798 by Count Rumford, an American who had fled to England in 1775 and eventually became a sort of military engineer to the Bavarian government. He became impressed by the large amount of heat that was developed in boring cannon, and he performed experiments indicating that this heat was too great to be accounted for plausibly by the caloric theory. He could find no loss of weight, when the chips made by boring were included, and showed that the specific heat of the chips was the same as that of the block from which they had come. He concluded that heat "cannot possibly be a material substance" such as caloric but must be a form of "motion."

A still more difficult experiment for the caloric theory to explain was one performed by Sir Humphry Davy, Director of the Royal Institution, which had been founded by Count Rumford. Davy rubbed together two pieces of ice in a vacuum surrounded by walls kept below the freezing point and found that the ice was melted. Here the mechanical work of rubbing accomplished exactly the same effects that could have been produced by the addition of a certain quantity of heat from the outside, yet there was no way in which caloric could have entered the ice.

The majority of the supporters of the caloric theory were, however, unconvinced. Even Carnot (1796-1832), the founder of the modern science of thermodynamics, when he proposed the now famous Carnot's cycle in 1824, based his reasoning on the caloric theory. A given

quantity of caloric "falling" from a higher to a lower temperature was analogous to a given quantity of water falling from a higher to a lower level; each was capable of producing motive power. The kinetic theory had to wait for a *quantitative* experiment.

In 1842, R. J. Mayer (1814–1878) published a paper¹ in which, partly on philosophical grounds, he announced the equivalence of heat and energy, and from data on the specific heats of a gas he deduced a value for the mechanical equivalent of heat. Meanwhile, Joule (1818–1889), in England, unacquainted with Mayer's work, was carrying on a very careful series of experiments in which he converted the mechanical energy of a falling weight into heat by a paddle wheel revolving in water and thus determined that 778 foot-pounds of work would raise 1 pound of water 1°F. Joule announced his results at a meeting of the British Association for the Advancement of Science in 1847. The paper would have passed almost unnoticed, had it not been for William Thomson, later Lord Kelvin, who, grasping the real significance of the proposed theory, by his discussion made the paper the event of the meeting.

Quite independently of the work of Mayer and of Joule, Helmholtz (1821–1894) in 1847 read a paper before the Physical Society in Berlin on "Die Erhaltung der Kraft," in which, on the basis of the impossibility of perpetual-motion machines, he announced the law of the conservation of energy. The paper was rejected for publication by the editor of the *Annalen der Physik*! It was later published in pamphlet form.

The caloric theory could not withstand these attacks, and by 1850 the mechanical theory of heat and the doctrine of the conservation of energy were generally accepted. The science of heat then proceeded to grow apace. The second law of thermodynamics was announced by Clausius (1850) and, in another form, by Kelvin (1851), and in 1854 Kelvin proposed the thermodynamic scale of temperature. Thus was developed the highly successful "classical theory" of heat. We shall discuss later the facts that this theory was unable to explain and the way in which the study of these facts led to the development of the quantum theory.

23. Light.—The revival of the wave theory of light, begun by Thomas Young (1773–1829), is one of the most important features of the history of the nineteenth century. Young pointed out that the dividing of a beam of light into a refracted ray at the interface between two mediums was to be expected from the wave theory but

¹ *Ann. d. Chem. u. Pharmacie*, May, 1842.

had not been satisfactorily explained on the corpuscular theory. In 1801, he presented to the Royal Society a paper "On the Theory of Light and Colors," in which he proposed the principle of the interference of two wave trains as an explanation of Newton's rings and the colors of thin plates. From Newton's measurements of the thickness of the air layers necessary to produce the several colors, Young was enabled to compute wave lengths. In subsequent papers, he described the interference fringes which he had observed by placing hairs or silk threads in front of a narrow slit illuminated from the rear; he announced the change of phase on reflection; he explained diffraction bands by the principle of interference, and he showed that the spacing of these bands gave values of the wave length agreeing with those obtained from Newton's rings and that, therefore, both phenomena must be due to a common cause. Again, *quantitative* measurements became an indispensable link in the chain of reasoning.

But the dogmatic spirit in regard to scientific matters was not yet dead. Young's paper aroused a storm of protest, even of derision and abuse. He was attacked not by the church, as was Galileo, but by some of his scientific, or, more probably, pseudoscientific, contemporaries. His chief assailant was Henry Brougham, afterward Lord Chancellor of England, who "reviewed" Young's papers in the *Edinburgh Review*. The nature of Brougham's attack is indicated by the following quotation:

We wish to raise our feeble voice against innovations that can have no other effect than to check the progress of science and renew all those wild phantoms of the imagination which Bacon and Newton put to flight from her temple. We wish to recall philosophers to the strict and severe methods of investigation . . .

Although Young replied at length in a privately published pamphlet, it was a long time before public opinion was willing to receive his theories with an open mind.

In 1815, a Frenchman, Fresnel (1788-1827), rediscovered the phenomenon of interference, performing the famous experiment with the two mirrors. A few years later he developed a mathematical theory of such phenomena (1818-1819). He also explained the polarization of light by assuming that the light vibrations in the ether are *transverse* to the direction of propagation of the light rather than longitudinal. He did not know that this suggestion had already been made by Young in a letter to Arago written in 1817. Fresnel supported the explanation by showing experimentally that two plane-polarized beams of light cannot interfere at all if their planes of

polarization are perpendicular to each other. Phenomena of polarization had been known to Newton, the polarization of light by Iceland spar being discovered by Bartholinus in 1669. Newton had tried to fit these phenomena into the corpuscular theory by assuming a sort of structure in the corpuscles, but the explanation was not convincing, and polarization had remained an enigma to both theories of light. Even Fresnel's explanation seemed almost to demand the impossible, for it required the ether to be a solid, or at least to have such properties of a solid as are necessary for the transmission of transverse waves, *i.e.*, the properties of rigidity and density. And yet the planets must move through this "solid" pervading all space, with no measurable changes in their periods of revolution!

Experimental evidence for the wave theory continued to accumulate. Finally, in 1850, Foucault performed a crucial experiment in its favor by showing, with his well-known revolving-mirror apparatus for measuring the velocity,¹ that light travels more slowly in water than in air, as it should on the wave theory, in order to account for the relative refractive index of these two mediums, whereas, with the corpuscular theory, the reverse should be the case.²

From 1850 until the end of the third period (1890), the wave theory held the field undisputed. The frequent assertions that the corpuscular theory was finally disposed of certainly seemed justified, particularly after the development of Maxwell's electromagnetic theory of light and its experimental verification. Yet the corpuscular theory was not dead. It was only sleeping.

Some important discoveries in light from 1800 to 1890 not previously mentioned are:

	DISCOVERER
Dark lines in the solar spectrum.....	Fraunhofer
Three-color theory of vision (1807).....	Young
Rotary polarization of quartz (1811).....	Arago
Polarization of scattered light (1813).....	Arago
Rotary polarization by liquids (1815).....	Biot
Light sensitivity of silver bromide (1826).....	Balard
Change of conductivity of selenium on illumination (1837).....	Knox
Doppler effect (1842).....	Doppler
Foundation of spectral analysis (1859).....	Kirchhoff and Bunsen

24. Electricity and Magnetism.—The history of electricity during the nineteenth century is so extensive that even a sketchy outline would fill a small volume. We shall, therefore, discuss little besides

¹ See EDSEB, "Light for Students."

² See any textbook on optics.

the fundamental discoveries of the opening decades and then the work of Faraday and that of Maxwell, which are so closely related to each other and to recent developments in physics that we can best present the part of electricity and magnetism in which we are particularly interested by giving an account of the contributions of these two men.

While the mathematical theory of electrostatics and of magnetism was being elaborated by Laplace, Green, and Poisson, fundamental discoveries were made in regard to electric *currents*. Galvani, in 1786, as a result of a chance observation that a frog's leg kicked convulsively when connected with the terminal of an electric machine, was led to an extensive study of "animal electricity." In the course of these experiments he observed that if the frog's leg was so suspended that the exposed nerves touched a metal plate, say silver, then, a contraction of the muscle occurred whenever the foot touched another metal, say iron. He even observed slight muscular contraction when both plates were of the same kind of metal. This led him to believe that the nerve was the source of electricity and that the metal served simply as conductor. Volta later found that potentials could be produced using inorganic materials and, in 1800, described the first battery for producing an electric current—the historically famous voltaic "pile," consisting of zinc and copper plates placed alternately and separated by blotting paper moistened with brine. He also described a battery consisting of cups containing brine or dilute acid connected by copper and zinc strips joined together.

Volta ascribed the effect to the contact of two dissimilar metals. We now know, however, that the current that can be caused in this manner to flow in a closed circuit is due to chemical action at the contacts of the metals with the electrolyte; the effect at the junction of the metals gives rise to potential differences of a different nature, the "contact difference of potential" or "Volta effect," which results in a total of zero in any closed circuit.

This new source of electricity was received with a great deal of interest. A few weeks after hearing of Volta's work, Nicholson and Carlisle accidentally discovered the decomposition of water by the electric current. Thinking to secure better contact between two wires forming part of the circuit, they had joined the ends of the wires by a drop of water. At once they observed the formation of a gas, which they recognized as hydrogen. This was the beginning of the study of electrolysis. During this same period the heating effect of the current and the arc light were discovered.

It was early suspected that there was some relation between electricity and magnetism, but the first significant discovery was made

in 1820 by Oersted, who found that a magnetic needle tends to set itself at right angles to a wire through which an electric current is flowing. Soon after, Biot and Savart discovered the law for the field of a long straight current, and, toward the end of 1820, Biot proposed, on the basis of a special experiment, the formula for the field due to a current element that is commonly used today and is often miscalled "Ampère's formula."¹ Soon after, the brilliant French physicist, Ampère (1775-1836), on the basis of just enough crucial experiments, showed that a closed current is equivalent in its magnetic effects to a magnetic shell. Then, reversing his line of thought, he suggested that magnetism itself might be due to currents circulating in the molecule. He also discovered the action of a magnetic field on a current. Thus, within five years of the first discovery, the foundations of electromagnetism had been completely laid.

25. Michael Faraday.² (a) *Biographical Sketch*.—Michael Faraday was born in 1791, in a small village near London. He was the son of a blacksmith, James Faraday. Being required to assist his mother in providing for the family, he was engaged in 1804 as errand boy to a bookseller and stationer, and in the following year he was formally apprenticed to his employer to learn the art of bookbinding. During this apprenticeship, Faraday made good use of his spare time by reading some of the books that passed through the shop. He was particularly interested in works on science, and in connection with his reading he showed one of the important characteristics³ of the great investigator-to-be by performing such of the simple experiments described "as could be defrayed in their expense by a few pence per week."

Aside from his own reading, *Faraday's only scientific education consisted in a dozen lectures on natural philosophy by a Mr. Tatum and four lectures on chemistry by Sir Humphry Davy, in the winter of 1812.* Submitting the very careful and neatly written notes which he made of these lectures "as proof of his earnestness," he made bold to apply to Sir Humphry Davy for a position, however menial, at the Royal Institution of which Davy was then director. Davy was so pleased with the letter and the notes that in March, 1813, Faraday was

¹ I.e., $dH = i ds \sin \theta / r^2$. Cf. Biot, J. B., "Précis élémentaire de Physique expérimentale," 1824. Ampère appears to have assumed that the force-action between two current elements must necessarily lie along a line drawn through the elements, which is not the case if Biot's formula is used for the magnetic field.

² See THOMPSON, SYLVANUS P., "Michael Faraday: His Life and Work."

³ In later life, he wrote: "I was never able to make a fact my own without seeing it."

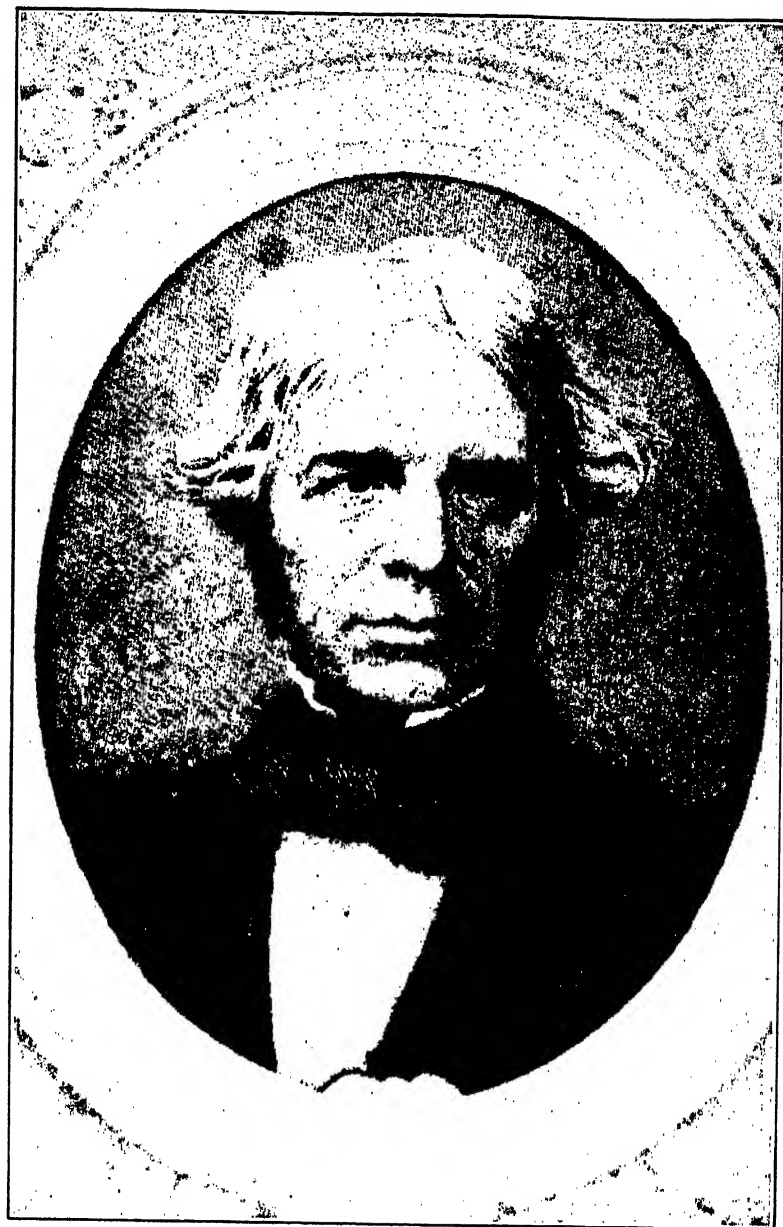


PLATE 3.—Faraday.

engaged as apparatus and lecture assistant at 25 shillings per week. In October, 1813, he accompanied Sir Humphry and Lady Davy on a trip to the Continent, which took them to many of the important scientific centers of Europe. Assistant though he was, Faraday impressed others because of his modesty, amiability, and intelligence; said one writer, "We admired Davy, we *loved* Faraday."

On returning to England, under Davy's encouragement, Faraday soon began original investigations, initially in chemistry. From 1816 to 1819 he published 37 papers. These were concerned with such subjects as the escape of gases through capillary tubes, the production of sound in tubes by flames, the combustion of the diamond, and the separation of manganese from iron. About 1820, he began his electrical researches. These, and others growing out of them, continued for nearly 40 years.

Almost his entire scientific life was spent at the Royal Institution. In 1825, he was made Director of the Laboratory. Declining offers of positions elsewhere, turning away professional occupations which might have made him wealthy, he gave to his science and to the institution he served a devotion seldom if ever equaled. The secret of his success, which brought him, during his lifetime, honors from all over the scientific world and which immortalized his name by the long list of scientific discoveries ascribed to him, is, perhaps, to be found in some excerpts from his many notes:

Aim at high things, but not presumptuously.

Endeavor to succeed—expect not to succeed.

It puzzles me greatly to know what makes the successful philosopher. Is it industry and perseverance with a moderate proportion of good sense and intelligence? Is not a modest assurance or earnestness a requisite? Do not many fail because they look rather to the renown to be acquired than to the pure acquisition of knowledge . . . ? I am sure I have seen many who would have been good and successful pursuers of science, and have gained themselves a high name, but that it was the name and the reward they were always looking forward to—the reward of the world's praise. In such there is always a shade of envy or regret over their minds and I cannot imagine a man making discoveries in science under these feelings.

The reader is urged to study carefully Faraday's life and works, particularly to read, as unexcelled examples of scientific expositions, portions of his "Experimental Researches in Electricity and Magnetism." We can mention here only a few of his most important discoveries.

(b) *The Principle of the Motor*.—Faraday had been interested in electromagnetism since April, 1821, when Wollaston attempted, at the

Royal Institution, to make a wire carrying an electric current revolve around its own axis when the pole of a magnet was brought near. The experiment was unsuccessful, but the phenomenon excited Faraday's interest, and he determined to make a study of it. First, he read what had been done by others and repeated many of their experiments. In the course of these experiments, he observed that, when the magnetic pole was brought near the wire, "the effort of the wire is always to pass off at right angles from the pole, indeed to go in a circle around it . . ."¹

The following day he wrote in his laboratory notebook:

Apparatus for revolution of wire and magnet. A deep basin with a bit of wax at bottom and then filled with mercury. A magnet stuck upright in wax so that pole [is] just above surface of mercury. Then piece of wire, floated by cork, at lower end dipping into mercury and above into silver cup.

On passing a current through the wire, it revolved *continuously* around the magnet. This was the first electric motor!

(c) *Electromagnetic Induction*.—Oersted's experiment and subsequent developments had clearly shown how "to produce magnetism by electricity." Faraday seems to have held it as one of the tenets of his scientific philosophy that every physical relation (of cause and effect) has its converse. If electricity can produce magnetism, then magnetism should produce electricity. His repeated attempts to accomplish this failed. For example, in 1825, he tried what seemed to be the obvious converse by looking for an electric current in a helix of wire coiled around a magnet. Later, he tried to find a current in a wire placed near another wire carrying current. Other scientists were looking for similar effects but without success. They were all looking for the production of a *steady* current.

But several times investigators were very near to the discovery of induced currents. In 1824, Arago observed the damping of the vibrations of a magnetic needle suspended over a copper plate. This observation was extended by causing the needle to revolve by revolving the copper plate underneath it, air disturbances being, of course, eliminated. It was shown that this "dragging" effect was greater, the greater the electrical conductivity of the spinning plate. Even the effect of radial slits in the copper disk, in reducing the dragging action on the magnet, was observed. Suggestive as these experiments were, however, the true explanation remained undiscovered.

In the summer of 1831 Faraday attacked the problem for a fifth time. This time, instead of placing a *permanent* magnet inside a

¹ Quotation from Faraday's laboratory notebook, Sept. 3, 1821.

helix, he procured a soft iron ring 6 inches in external diameter, on which he wound two coils of copper, *A* and *B*, "separated by twine and calico." To detect a possible current in coil *B*, he "connected its extremities by a copper wire passing to a distance and just over a magnetic needle." When coil *A* was connected to a battery, there was "a sensible effect on the needle. It *oscillated* and settled at last in *original position*. On breaking connection of side *A* with battery, again a disturbance of the needle." Slight as these momentary effects were, Faraday recognized their importance, although he too had been looking for a *continuous* effect. On Aug. 30, he writes, "May not these transient effects be connected with causes of difference between power of metals at rest and in motion in Arago's experiments?"

From this slender clue, Faraday proceeded rapidly to the discovery of the real effect. On the "third day" of his experiments, he wound a coil of wire around an iron cylinder and placed the cylinder so as to join the *N* pole of one permanent magnet with the *S* pole of another. The coil was connected to a galvanometer:

Every time the magnetic contact at *N* or *S* was made or broken there was a magnetic action at the indicating helix [*i.e.*, galvanometer]—the effect being, as in former cases, not permanent but a mere momentary push or pull.

On the fourth day, he showed that the presence of iron was not necessary; that the effect could be produced by the action of one helix on another. On the fifth day:

A cylindrical bar magnet . . . had one end just inserted into the end of the helix cylinder; then it was quickly thrust in the whole length and the galvanometer needle moved; then pulled out and again the needle moved, but in the opposite direction. The effect was repeated every time the magnet was put in or out, and therefore a wave of electricity was so produced from mere *approximation* of a magnet and not from its formation *in situ*.

At last! He had "converted magnetism into electricity." The essential requisite was *relative motion*, or a *change* of condition. On the ninth day, he produced a continuous current by turning a copper disk between the poles of a powerful electromagnet, the periphery of the disk being connected to its axis through an indicating galvanometer. This was the now well-known Faraday disk dynamo, the *very first* dynamoelectric machine.

Thus, after only a few days' work in his laboratory, following, however, years of patient and persistent experiment, Faraday had discovered a phenomenon for which the greatest scientists of his time had sought in vain—electromagnetic induction.

Following this discovery, Faraday devised and tried various electric machines to test and extend his newly discovered principle. One of these machines, consisting of a rotating rectangle of wire *with a commutator attached*, is the prototype of the modern dynamo. But his interest was always in pure science, for he writes:

I have rather, however, been desirous of discovering new facts and relations dependent on magnetoelectric induction, than of exalting the force of those already obtained; being assured that the latter would find their full development hereafter.

Being unacquainted with mathematical symbols and methods, Faraday always sought to explain his discoveries and to extend his researches by purely physical reasoning. To the mathematician, the law of magnetic attraction

$$F = \frac{m_1 m_2}{r^2}$$

may have been a sufficient explanation of the phenomenon. To Faraday, this gave a statement only of the *magnitude* of the magnetic forces; *it left the phenomenon itself quite unexplained*. Accordingly, he insisted that two magnetic poles, or two electric charges, could act on each other *only if the medium between the two played some important part in the phenomenon*. This insistence on the importance of the medium ultimately led him to the very fruitful concept of lines of force and of the "cutting" of these lines as essential to electromagnetic induction. At first qualitative, this concept was developed by Faraday into an essentially quantitative form, although it was first stated in mathematical language by F. Neumann in 1845. Commenting on Faraday's laws of electromagnetic induction, Maxwell wrote:

After nearly a half-century . . . , we may say that, though the practical applications of Faraday's discoveries have increased and are increasing in number and value every year, no exception to the statement of these laws as given by Faraday has been discovered, no new law has been added to them, and Faraday's original statement remains to this day the only one which asserts no more than can be verified by experiment, and the only one by which the theory of the phenomena can be expressed in a manner which is exactly and numerically accurate, and at the same time within the range of elementary methods of exposition.

(d) *The Laws of Electrolysis*.—Faraday next turned his attention to proving that "Electricity, whatever may be its source, is identical in its nature." He found, for example, that electricity from a friction machine would deflect a galvanometer and would cause chemical

decomposition just as would electricity produced by chemical action. This led him into the field of electrolysis. He found that many substances, such as certain chlorides and sulphates, are nonconductors when solid but are good conductors when melted, and that in the molten state they are decomposed by the passage of current. This showed that water was *not* essential to electrolysis. To clarify description of his experiments, he introduced the terms "electrode," "anode," "cathode," "ion," "anion," "cation," "electrolyte," "electrochemical equivalent," etc. A quantitative study of the phenomena resulted in his discovery of the laws of electrolysis that bear his name and which are the basis of all present-day work in that field.

Further, Faraday clearly recognized that a definite quantity of electricity is associated with each atom or ion in electrolysis. Had he been able to determine the number of atoms in unit mass of any substance, he would have anticipated, by 60 years, the determination of the fundamental charge e . For he says:

Equivalent weights of bodies are simply those quantities of them which contain equal quantities of electricity; . . . it being the *electricity* which determines the combining force. Or, if we adopt the atomic theory or phraseology, then the atoms of bodies which are equivalent to each other in their ordinary chemical action, have equal quantities of electricity naturally associated with them.

(c) *The Conservation of Energy*.—In connection with a proof of the fact that the electricity from the voltaic pile results from chemical action and not from mere contact of one substance with another, Faraday stated clearly the doctrine of the conservation of energy several years before the statement of Helmholtz. In 1840, he wrote:

The contact theory assumes that a force which is able to overcome a powerful resistance . . . can arise out of nothing . . . This would indeed be a creation of power, and is like no other force in nature. We have many processes by which the form of the power is so changed that an apparent conversion of one into the other takes place . . . But in no case is there a pure creation or a production of power *without a corresponding exhaustion of something to supply it*.

(f) *The "Faraday" Effect*.—Reference has already been made to Faraday's abhorrence of the doctrine of "action at a distance." He believed that, if two electric charges attract each other, the medium between the two plays some important role. Presumably, therefore, the medium between two such charges is in a different state than it would be if the charges were not present; and if so, such an altered state should be detectable by observing the alteration in some physical

property of the medium. As early as 1822, Faraday had experimented with a beam of polarized light passing through a transparent solution carrying a current, to see whether the current caused any "depolarizing" action. Although he repeated the experiment several times in subsequent years, the results were all negative. In 1845, he returned to the problem, but still with negative results. He then tried solid dielectrics between plates of metal foil connected to a powerful electric machine to see whether, under electric strains, they would show any optical effects. No results!¹

Faraday then substituted a magnetic field for the electrostatic field to see whether the former would cause any depolarizing action on the beam of light. Various substances were tried but still with negative results. Finally, he placed in the magnetic field a very dense piece of lead glass, which he had made many years earlier. When the magnetic lines were parallel to the direction of the beam of polarized light, he observed that the plane of polarization was rotated. At last, he had found a relation between magnetism and light. This magnetic rotation is now known as the "Faraday effect." Again, his persistent search, maintained during 20 years of repeated failures, was rewarded by the discovery of an effect in the existence of which he had the most sublime confidence.

(g) *Miscellaneous*.—Among Faraday's other researches may be mentioned: numerous investigations in chemistry; the liquefaction of several gases formerly thought "permanent"; the diffusion of gases through solids; self-induction; certain fundamental properties of dielectrics; diamagnetism; distinction between anode and cathode in the electric discharge through gases at low pressure; vibration of plates; regelation of ice; alloys of steel; and optical glass.

Well may this simple, modest, self-taught philosopher be given a conspicuous place among the great benefactors of mankind.

26. Joseph Henry (1799-1878).—Any account of Faraday's work, however brief, should be accompanied by at least a mention of the researches of the American physicist, Joseph Henry, whose memory is honored by the name of the unit of inductance, the henry, which bears to electrokinetics a relation identical with that of the farad to electrostatics. Had Henry been able to experiment continuously, and with more resources, instead of only during a summer vacation of 1 month while teaching mathematics at Albany Academy, and then only with such apparatus as he could make with his own hands, he would undoubtedly have anticipated Faraday in the discovery of electro-

¹ Years later (1875) this effect was found by Kerr [*Phil. Mag.*, vol. 1, p. 337 (1875)].

magnetic induction, including the phenomena of self-induction. In all of his work, furthermore, he was greatly hampered by his isolation from the scientific atmosphere of Europe.

Henry was interested especially in the design and use of electromagnets. He constructed the first electric motor operating by an electromagnet, which rocked back and forth between two permanent magnets. He found that, for maximum tractive effect, the cells of the battery and also the "spools" of the electromagnet should be connected in series if the magnet were a long distance from the battery, but they should be connected in parallel if the wires joining the magnet to the battery were short. His work on electromagnets led *directly* to the commercial development of the telegraph.

27. James Clerk Maxwell (1831-1879).—It would be difficult to pick out two eminent scientists whose beginnings differed from each other more than did Maxwell's and Faraday's. Faraday came of very humble parentage; Maxwell, from a long line of distinguished ancestors. Faraday's early life was lived almost in poverty; Maxwell's family had abundant means. Faraday received only the most rudimentary education; Maxwell was given every advantage of school and university. They differed also in their aptitude for scientific work. Faraday was one of the greatest exponents of experimental science that the world has ever seen; whereas Maxwell, although likewise an able experimenter, is one of the greatest figures in the history of theoretical physics. And yet both made indispensable and mutually supplementary contributions to the classical theory of electromagnetics.

Maxwell was born in Edinburgh in 1831. At the age of ten he was sent to the Edinburgh Academy, where he was a friendly boy, though never quite amalgamating with the rest. But, however strange he sometimes seemed to his companions, he had three qualities which they could not fail to understand: agile strength of limb, imperturbable courage, and profound good nature.¹

When he left the Academy in 1847, he was "first in mathematics and in English and nearly first in Latin." Then, after 3 years in the University of Edinburgh, he entered Trinity College, Cambridge, from which he graduated in 1854 with high honors.

Maxwell early showed extraordinary interest in both theoretical and experimental research in physics. At the academy, he invented a means of drawing certain types of oval curves, and a few years later he published a paper on "The Theory of Rolling Curves" and another on "The Equilibrium of Elastic Solids"—all this before he was nineteen years old! During these same years he was also busy with

¹ GLAZEBROOK, "James Clerk Maxwell and Modern Physics."



PLATE 4.—Maxwell.

experiments of many sorts, especially in his little laboratory in a garret on the family estate at Glenlair, where he spent his vacations.

After 4 years at Aberdeen, he was Professor for 5 years at King's College, London (1860-1865); from here some of his most important papers were published, such as "Physical Lines of Force" (1862) and his greatest paper, "A Dynamical Theory of the Electromagnetic Field." After a retirement of several years, he was elected in 1870 to the newly founded professorship of experimental physics at Cambridge. In this capacity, he superintended the planning and equipment of the now famous Cavendish Laboratory, of which he was director until his untimely death in 1879.

A large proportion of Maxwell's papers, over 100 in number, may be grouped under three headings: color vision; molecular theory; and electromagnetic theory.

The work on color vision was undertaken to make a quantitative study of the physical facts pertinent to the theory of color sensations proposed by Thomas Young, according to which any luminous sensation is the result of exciting in the eye three primary sensations, red, green, and violet. For this purpose Maxwell invented a "color box," by means of which he could mix spectral colors.

Maxwell's work on molecular physics is very extensive. He discovered and, in part, established theoretically the law of the distribution of velocities among the molecules of a gas ("Maxwell's law"). He showed that when two gases are at the same temperature, the mean kinetic energy of translatory motion of their individual molecules is the same in both gases. From the kinetic theory of viscosity, he drew the surprising conclusion that the viscosity of a gas should be independent of its density so long as the mean free path is not too large, and he verified this conclusion by experiment. He brought to bear upon the whole subject mathematical methods "far in advance of anything previously attempted on the subject"; indeed, he is the co-founder with Clausius (1822-1888) of the kinetic theory of matter.

In the electromagnetic theory, Maxwell's great contributions were the "displacement currents" and the formulation of the general equations of the electromagnetic field, which led to the electromagnetic theory of light. In the preface to his treatise "Electricity and Magnetism," he makes the interesting remark:

Before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's "Experimental Researches on Electricity."

He became convinced that Faraday was right in regarding the dielectric as the true seat of electrical phenomena and in supposing that it acted by becoming electrically polarized, the positive ends of its molecules pointing on the whole with the field and the negative ends in the opposite direction. The term "dielectric," as used here, must be understood to include a tenuous medium or ether filling all space, even in what we call a vacuum. He drew the conclusion that when the polarization changes, this change must involve a displacement of electricity, and so there must exist in the dielectric, while the change is going on, a current having the same magnetic properties as the current in a conductor.

This assumption of displacement currents opened the way for the deduction of Maxwell's famous equations of the electromagnetic field. It is interesting, however, that he was first led to these equations through a mechanical analogy, *i.e.*, in studying the behavior of a *mechanical system* filling all space, which would be capable of causing the observed electrical and magnetic phenomena. He showed (1862) that his hypothetical medium would be capable of transmitting transverse vibrations with a speed equal to the ratio of the electromagnetic to the electrostatic unit of charge. Although he did not take his model too seriously, he nevertheless remarks, that the ratio of the units

. . . agrees so exactly with the velocity of light calculated from the optical experiments of M. Fizeau, that we can scarcely avoid the inference that *light consists in the transverse undulations of the same medium, which is the cause of electric and magnetic phenomena.*

The theory was restated, without reference to any particular model, in his great paper of 1864, in which he says:

The theory which I propose may therefore be called a theory of the *Electromagnetic Field*, because it has to do with the space in the neighborhood of the electric or magnetic bodies, and it may be called a *Dynamical Theory*, because it assumes that in that space there is matter in motion by which the observed phenomena are produced.

In 1873, Maxwell published his "Treatise on Electricity and Magnetism," which ranks with Newton's "Principia" as one of the most important books in all science.

Maxwell's equations and the subject of electromagnetic waves will be discussed in some detail in the next chapter along with certain other results of electromagnetic theory that will be needed later. Here we shall add only a few words concerning the later history of the subject during the closing decades of the last century.

28. The Completion of Electromagnetic Theory.—The *physical ideas* underlying Maxwell's new theory were left none too clear by him. In his treatise, we find the assumption that all space is full of incompressible "electricity"; in a conductor this electricity can move freely (except for ohmic resistance), thus constituting an electric current, but in a dielectric "there is a force which we have called electric elasticity which acts against the electric displacement and forces the electricity back when the electromotive force is removed." This is clear enough. But what is the origin of this "electromotive force?" And in what does electrification consist? According to Maxwell's theory a charged conductor has neither more nor less electricity on it or in its neighborhood than when it is not charged. This point became still more obscure when in 1876 the American physicist, Rowland, showed experimentally that a moving charged conductor is surrounded by a magnetic field, the moving electrification evidently constituting a current.

The *mathematical theory*, on the other hand, was slowly developed by others, especially by H. A. Lorentz, and was shown to give a good account of all electric and magnetic phenomena and of the principal properties of light. And in Germany, stimulated by Helmholtz, Hertz set out to search experimentally for the magnetic effects of Maxwell's displacement currents and in 1887 discovered waves that were undoubtedly of electrical nature. Later, it was shown that the speed of propagation of these waves is the same as that of light. Speculation as to the nature of the displacement currents in a vacuum then gradually died out, until today we speak only of electric and magnetic "fields" governed by the Maxwell-Lorentz equations. This development in electromagnetic theory illustrates a general tendency, notable during the last century but regretted by many, for the fundamentals of physics to become an abstract mathematical theory unsupported by underlying concrete ideas.

As we have noted in the Introduction, about 1880 a few physicists were thinking that perhaps all of the *important* discoveries in physics had been made, so that research henceforth would consist in studying details and in improving the technique of measurement "so as to investigate the next decimal place." We now know that in reality more fundamental discoveries yet were to come. It will always be unsafe to assume that all of the important problems in physics have been solved.

CHAPTER II

ELECTROMAGNETIC WAVES AND MOVING CHARGES

Before taking up the main theme of this book, we shall discuss, in this chapter, a few points in electromagnetic theory which it will be useful for the student to have in mind. Electromagnetic waves and the emission of radiation will be discussed, on the basis of Maxwell's theory, in sufficient detail to enable the student to appreciate more readily the conflict of this theory with the modern quantum theory. Then a few points concerning the fields of moving charges will be taken up. We begin with a detailed discussion of Maxwell's equations for the electromagnetic field.

29. Displacement Currents.—In discussing Maxwell's new theory of electromagnetism, we have seen that his greatest contribution was the displacement current. This is a current which, capable of existence even in a vacuum, is associated with the magnetic field in the same way as is an ordinary current in a wire and which is of such magnitude that all currents are effectively closed. To find what the necessary magnitude of the displacement current is, consider the circuit shown in Fig. 1. A battery B is connected to a long resistance ab on which is a sliding contact c . Points a and c are connected through galvanometers G_1 and G_2 to the plates of a condenser C . If c , starting from coincidence with a , moves uniformly toward b so as to increase uniformly the potential difference between the plates of the condenser, thereby increasing its charge, a constant current, say i , will flow through the galvanometer G_1 and an identical current through G_2 . Except for the condenser, the circuit aG_1G_2c might be regarded as a continuous circuit. No actual current, in the ordinary sense of the term, flows between the plates of the condenser; but, during the charging process, the *electric field* between the plate is *increasing* at a uniform rate.

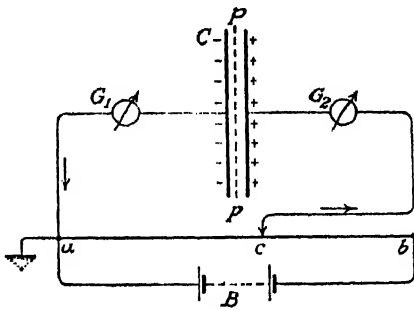


FIG. 1.

merely assumed that, wherever an electric field is changing, there is also a magnetic field related to the displacement current in exactly the same way as the magnetic field is related to a conduction current having the same direction and magnitude. (A less sophisticated form of statement would be that displacement currents "produce" magnetic fields according to the same law according to which such fields are produced by conduction currents. It is not really obvious, however, which is cause and which is effect. Does the displacement current cause the magnetic field, or do certain peculiarities in the geometrical distribution of the magnetic field cause the change in the electric field, in accordance with Maxwell's equations?)

30. Maxwell's Equations.—We are now in a position to deduce Maxwell's equations for the electromagnetic field. In order to have these equations collected together in a group, we shall first write them all down, and then we shall proceed to consider their deduction from the familiar laws of electromagnetism.

The electric and the magnetic field strengths are vectors, having direction as well as magnitude. To deal with them analytically, let us introduce a set of right-handed Cartesian axes. To save words, it will be convenient, hereafter, to say that any three mutually perpendicular vectors related to each other in direction like the directed lines Ox , Oy , Oz form a *right-handed (orthogonal) set*¹ of vectors.

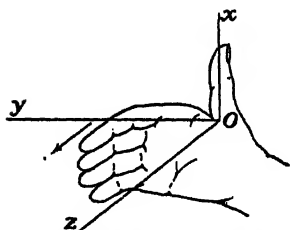


FIG. 4.—Illustrating the right-hand vector rule.

Let us now represent the field vectors by their components in the directions of these axes. Let the components of the electric field strength E be denoted by E_x , E_y , E_z , respectively, and those of the magnetic field strength H by H_x , H_y , H_z . Then Max-

well's field equations for a region in which the dielectric constant is ϵ and the magnetic permeability μ , but in which there is no electrical conductivity and no accumulation of electric charge, can be written thus:

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \frac{\epsilon}{c} \frac{\partial E_x}{\partial t}, \quad (3a)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \frac{\epsilon}{c} \frac{\partial E_y}{\partial t}, \quad (3b)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \frac{\epsilon}{c} \frac{\partial E_z}{\partial t}, \quad (3c)$$

¹ Vectors can be tested for this property with the right hand, as suggested in Fig. 4.

$$\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\mu}{c} \frac{\partial H_z}{\partial t}, \quad (3d)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\mu}{c} \frac{\partial H_y}{\partial t}, \quad (3e)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\mu}{c} \frac{\partial H_z}{\partial t}, \quad (3f)$$

$$\frac{\partial}{\partial x}(\mu H_x) + \frac{\partial}{\partial y}(\mu H_y) + \frac{\partial}{\partial z}(\mu H_z) = 0, \quad (3g)$$

$$\frac{\partial}{\partial x}(\epsilon E_x) + \frac{\partial}{\partial y}(\epsilon E_y) + \frac{\partial}{\partial z}(\epsilon E_z) = 0. \quad (3h)$$

We shall suppose that c.g.s. units are used and that E is in electrostatic units, whereas H is in electromagnetic units or oersteds; ϵ and μ are, therefore, pure numbers, and equal to unity if the field is in a vacuum. The derivatives occurring in the equations are all partial derivatives, since $E_x, E_y, E_z, H_x, H_y, H_z$ may all vary, not only from point to point, but also from time to time, each component being thus a function of the four variables x, y, z, t . The constant c enters into these equations as the ratio of the electromagnetic to the electrostatic unit of charge; as is well known, however, this constant is also equal to the speed of light in vacuum.

The mathematical foundation of the theory may then be completed by adding to the field equations the Lorentz force equation, which merely expresses the familiar value for the force exerted by the field upon a small test charge placed in it. Upon a stationary charge of q electrostatic units, there acts a force of qE dynes in the direction of E ; if the charge is moving with velocity v cm. sec.⁻¹, there is added to the force qE another force of magnitude $qvH \sin \theta/c$ dynes, θ being the angle between the directions of v and of H . This additional force acts perpendicularly both to v and to H and in such a direction that the three vectors \mathbf{v} , vector component of H perpendicular to \mathbf{v} , and the force itself form a right-handed set (see above and Fig. 4).

Of the six Maxwell equations, number (3h) is a mathematical equivalent of Gauss' theorem, familiar in electrostatics. Equation (3g) represents the corresponding fact in the theory of magnetism, i.e., that the tubes of the induction μH are closed, each tube having a uniform flux of induction across all of its cross sections. We shall not actually make use of these equations ourselves, preferring to employ the corresponding physical laws directly; hence we shall refer for their proof to the standard treatises on Electricity and Magnetism.¹ Proofs

¹ E.g., STARLING, "Electricity and Magnetism," Chap. XIII (or XIV); PAGE and ADAMS, "Principles of Electricity," Chap. XVI.

of the remaining equations are also given in those treatises, but it may be worth while to repeat the proof of these equations here.

Equations (3a) to (3c) are a mathematical equivalent of the familiar principle that the work done by the field on a unit positive magnetic pole carried around any closed path equals 4π times the total current, in electromagnetic units, that is encircled or linked by the path. (The principle can also be stated more elegantly in terms of a line integral without reference to poles, which do not exist, but the statement in terms of a fictitious pole has the advantage of concreteness.) In the present instance, however, we are dealing with a medium in which only displacement currents can occur. It may be remarked that the principle just stated could not be formulated as a general one, applicable to all cases, if displacement currents did not exist; for then the currents, being limited to those on conductors, might not be closed, and in such cases the expression "current linked by a path" would not convey a precisely defined meaning. For this reason, it was impossible for any one to arrive at Maxwell's field equations until displacement currents, or their equivalent, had been invented.

To obtain equation (3a) from the principle just stated, draw a small rectangle $PQRS$ with one corner at any chosen point P ; let it lie in a plane perpendicular to the x -axis and have two sides of length dy parallel to the y -axis and the other two of length dz parallel to the z -axis (cf. Fig. 5). Let a unit magnetic pole be carried in the direction $PQRS$ once around this rectangle. Then, if we let H_y denote the value of the y -component of H on QR and H'_y its value on the opposite side SP , treating these values as constant along their

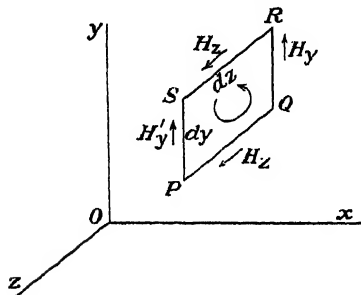


FIG. 5.

respective sides because the rectangle is small, the work done by the field on the pole, as it traverses these two sides, is $H_y dy - H'_y dy$. Treating the other two sides similarly, we have, for the total work on the pole,

$$dW = (H'_z - H_z) dz - (H'_y - H_y) dy. \quad (4)$$

But two points situated like P and Q , or S and R , differ only in their z -coordinates; hence the ratio $(H'_y - H_y)/dz$ will be close to the partial derivative of H_y with respect to z at P ; this ratio will approach the

derivative as a limit as $dz \rightarrow 0$.^{*} We can write, therefore,

$$H'_y - H_y = \frac{\partial H_y}{\partial z} dz.$$

Treating $H'_z - H_z$ in the same way and then substituting in (4), we have

$$dW = \frac{\partial H_z}{\partial y} dydz - \frac{\partial H_y}{\partial z} dzdy. \quad (5)$$

On the other hand, the total amount of current linking with the path $PQRS$, or the current flowing through the rectangle, is determined by the x -component of the displacement current. It may help the student to visualize the situation if the rectangle is supposed drawn in a large tank of electrolytic solution with current entering at one point and leaving at another point. If the current flows perpendicularly across the rectangle with a density j_x , the total current crossing it is $j_x dydz$. If the electricity has additional components of motion parallel to the rectangle, these do not affect the amount of electricity that crosses the rectangle. So here, from expression (2) for the density of displacement current j_ν , we have for the total displacement current linking with the path

$$j_\nu dydz = \frac{\epsilon}{4\pi} \frac{\partial E_x}{\partial t} dydz.$$

Upon multiplying this expression for the current by 4π and dividing it by c , the ratio of the electromagnetic to the electrostatic unit of charge, in order to convert the current into electromagnetic units, and then equating the result to the expression found above for the work dW on a unit pole, we have

$$\frac{\epsilon}{c} \frac{\partial E_x}{\partial t} dydz = \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \right) dydz.$$

This gives Eq. (3a) upon dividing through by $dydz$.

Equations (3b) and (3c) can be derived in a similar manner by using rectangles perpendicular to the y - and z -axes, respectively.

^{*} In case the reader may wonder why the variation of H_y and H'_y along their respective sides of the rectangle does not have to be taken into consideration, as well as their variation in the perpendicular direction, it may be remarked that H_y and H'_y will vary in approximately the same manner along QR and along PS , respectively, and the difference of these variations gives rise to a term in dW of a higher order, which can be neglected. The argument can be made entirely rigorous, provided $\partial H_y / \partial z$ is a continuous function of y and z .

It is easier, however, to obtain them by *cyclic interchange* of the variables. We might have drawn our x -axis in the direction Oy in Fig. 5 and our y -axis in the direction Oz ; then, to keep the axes right-handed, the z -axis would have the direction Ox , for a rotation of the axes like a rigid body through 90° about the old Oz would bring the x -axis onto Oy , the y -axis now falling on $-Ox$. Then a second rotation of 90° about the old Oy would bring the y -axis onto the old Oz and at the same time the z -axis would be brought onto the old Ox . Equation (3a) can now be deduced in terms of the new variables by the same argument as before. But then, since it does not matter mathematically what letter is used for a variable, we can replace the new variables by the old ones in the end, which amounts to making the *cyclic interchange*:

$$x \rightarrow y, \quad y \rightarrow z, \quad z \rightarrow x. \quad (6)$$

Thus we obtain (3b). Another cyclic interchange then gives us (3c).

The remaining three equations (3d) to (3f) embody in a similar way Faraday's principle that the electromotive force in a closed circuit, measured in electromagnetic units, equals the negative of the time rate of change of the flux of magnetic induction through the circuit. The electromotive force is measured by the work done on a unit positive charge carried around the circuit. Using the same rectangle as before (Fig. 5), we find, by the same argument as that which led to Eq. (5), for the work done on an *electromagnetic* unit charge, which is equal to c electrostatic units, when it is carried around the rectangle:

$$dW = c \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) dydz.$$

The magnetic flux through the rectangle is

$$\mu H_x dydz.$$

For we can think of the field as made up of three fields superposed, each parallel to one of the axes, and of magnitudes H_x , H_y , H_z , respectively; then it is clear that the H_y and H_z fields contribute nothing to the flux through the rectangle, being parallel to its plane. The negative rate of change of this flux is, therefore,

$$-\mu \frac{\partial H_x}{\partial t} dydz,$$

since μ and $dydz$ do not change with time. Equating this to the expression just found for dW and dividing through by $c dydz$, we have Eq. (3d). Successive cyclic interchanges then give (3e) and (3f).

It might be remarked, by way of generalization, that if electric charges are present in the field, with a volume density ρ in electrostatic units, the 0 in Eq. (3h) is replaced by $4\pi\rho$; and, if currents are present, with density components j_x, j_y, j_z in electrostatic units (i.e., j_x = amount of electricity flowing per second across a unit area placed perpendicular to the x -axis, at any point in the medium), there are added to the right-hand members of (3a) to (3c) the terms $4\pi j_x/c, 4\pi j_y/c, 4\pi j_z/c$, respectively.

31. Energy and Momentum in the Electromagnetic Field.—In the interaction of material bodies with electromagnetic fields, the bodies commonly gain or lose *energy*. It has been found possible to preserve the law of the conservation of energy by supposing a definite amount of energy to exist in the field as well as in material bodies. This is the usual procedure in physics. Whenever a new phenomenon is encountered, the attempt is made to invent a new form of energy so that the conservation law will hold; and so far it has always proved possible to do this.

The energy in the field is determined by calculating the work done in setting it up. In this way it can be proved quite generally,¹ and is shown for special cases in elementary textbooks, that energy can be supposed to be distributed throughout an electromagnetic field at the rate of

$$w = \frac{\epsilon E^2 + \mu H^2}{8\pi} \text{ ergs per unit volume,} \quad (7)$$

E being the electric intensity in electrostatic units, H the magnetic intensity in oersteds, ϵ the dielectric constant, and μ the permeability.²

When the field changes, the energy in it must be supposed to move about. From the theory, it can be shown that correct results are obtained if we suppose the energy to stream in the direction of a vector known as *Poynting's vector*. This is a vector drawn perpendicular to E and to H and in such a direction that a right-handed set of vectors is formed by the vector E , the vector component of H perpendicular to E , and Poynting's vector (cf. Sec. 30 and Fig. 6). Its magnitude is assumed to be

$$\frac{c}{4\pi} EH \sin \theta$$

¹ Cf. SMYTHE, W. R., "Static and Dynamic Electricity," Sec. 13.04.

² This expression is restricted, of course, to free space or to matter that is isotropic and not ferromagnetic. In material bodies, furthermore, w represents free energy, in the thermodynamic sense, rather than stored energy; but this distinction is seldom important.

where θ denotes the angle between the directions of E and H , and c is the ratio of the electromagnetic to the electrostatic unit of charge, or the speed of light in vacuum.

This last expression represents the rate at which the energy is flowing. If, at any point P in the field, a small plane surface is drawn parallel to E and to H and hence perpendicular to Poynting's vector (cf. Fig. 6), then the number of ergs of energy crossing each square centimeter of this area per second is given by the magnitude of Poynting's vector as just written. The student can readily verify in simple cases the fact that the correct amount of energy is thereby delivered to the various parts of electrical circuits; e.g., the heat produced in a resistor by an electric current can be supposed to stream in from the surrounding space in the direction of Poynting's vector.

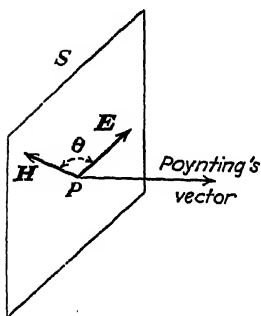


FIG. 6.

In mechanics, the *principles of momentum* are second only to those of energy in importance. Now electromagnetic fields are capable of changing the momentum of material bodies. In all practical cases the *total momentum* of all bodies concerned remains constant, a force on one body being offset by a reaction on some other body. There are a few phenomena in which this is not true, however; the most noteworthy is the phenomenon of light pressure, discussed below. It can be shown¹ that the principles of momentum and of angular momentum can be preserved even in such cases provided we assign momentum not only to matter but also to the electromagnetic field itself.

The momentum in the field is called *electromagnetic momentum*. It is a vector quantity, of course, and it is assumed to have, at each point in the field, the direction of Poynting's vector at that point; furthermore, the amount of electromagnetic momentum in unit volume, in space free from matter, is just equal to Poynting's vector divided by c^2 .

Space cannot be taken here to deduce either of these two principles from the fundamental electromagnetic equations. It can easily be seen that they can be interpreted in a concrete way that is of great interest in connection with modern ideas concerning mass and energy. We can, in fact, account for the transfer of energy by supposing that the energy in the field whose density is w is actually in motion with a

¹ Cf. ABRAHAM, M., and R. BECKER, "Classical Electricity and Magnetism," Chap. XV; also LORENTZ, H. A., "Theory of Electrons," Secs. 19-25 (1916), a readable discussion.

certain velocity v in the direction of Poynting's vector Π . The necessary velocity is

$$v = \frac{\Pi}{w};$$

for, if we draw a right cylinder of height v and of unit cross section, with its length in the direction of Poynting's vector, this cylinder will contain wv or Π ergs of energy; and in a second this energy moves across the end of the cylinder, thus producing the rate of transfer required by Poynting's vector. If we then add the assumption that *each erg of the moving energy has a mass of $1/c^2$ or $1/(9 \times 10^{20})$ gram*, the energy in a cubic centimeter will have momentum equal to

$$(w/c^2)v = \Pi/c^2,$$

which, in free space, is the magnitude of the momentum in unit volume. Thus, the electromagnetic momentum in free space may be thought of as ordinary momentum possessed by the moving electromagnetic energy.

32. Electromagnetic Waves.—The field equations [(3a) to (3h)] admit of an enormous variety of different solutions. There is no general method for discovering useful solutions; in the end, one has to resort to guessing. Accordingly, a particular solution that turns out to be important will simply be written down here and then shown to satisfy the equations. Suppose that

$$E_y = f(x - vt), \quad E_x = E_z = 0, \quad (8)$$

$$H_z = \sqrt{\frac{\epsilon}{\mu}} f(x - vt), \quad H_x = H_y = 0, \quad v = \frac{c}{\sqrt{\epsilon\mu}}. \quad (9)$$

Here f stands for any function whatever (provided it has a first derivative). By substitution it is easily seen that all eight of the field equations are satisfied. Substitution in (3b), for example, gives

$$-\sqrt{\frac{\epsilon}{\mu}} \frac{\partial f}{\partial x} = \frac{\epsilon}{c} \frac{\partial f}{\partial t}. \quad (10)$$

But

$$\frac{\partial f}{\partial x} = \frac{df}{d(x - vt)} \frac{\partial(x - vt)}{\partial x} = \frac{df}{d(x - vt)} \times 1, \quad (11)$$

$$\frac{\partial f}{\partial t} = \frac{df}{d(x - vt)} \frac{\partial(x - vt)}{\partial t} = -v \frac{df}{d(x - vt)} = -\frac{c}{\sqrt{\epsilon\mu}} \frac{df}{d(x - vt)} \quad (12)$$

because of the value assumed for v . Substituting these expressions for the derivatives in (10), we obtain an identity.

As an alternative to using Eqs. (3g) and (3h), which we did not deduce above, we may simply note that according to our assumed solution the electric lines are parallel to the y -axis and therefore straight, and the field strength is also uniform, so that Gauss' theorem is satisfied. A corresponding remark can be made about the magnetic field.

The physical characteristics of the field specified by this solution are simple and interesting. It represents *waves* of electric and magnetic field moving toward $+x$ with *speed* v . For, if a point is made to move toward $+x$ with a speed equal to v , at this point the quantity $x - vt$ always keeps the same value, since, in time Δt , x changes by $\Delta x = v \Delta t$, and $x - vt$ changes, therefore, by

$$\Delta(x - vt) = \Delta x - v \Delta t = 0.$$

Hence, at this moving point, E_y and H_z always keep the same values. Furthermore, the waves are *plane*, since E_y has the same value at all points of any plane perpendicular to the direction of propagation, which is the x -axis; and so has H_z , also. Finally, the waves are *transverse*, for the vectors E and H are both perpendicular to the direction of propagation; and the waves are *plane-polarized*, with the electric vector in the xy plane.

As a special case, the function f might be sinusoidal in form. For example, with x' standing for any variable, we might take

$$f(x') = A \sin(2\pi x'/\lambda).$$

Then

$$f(x - vt) = A \sin 2\pi \left(\frac{x}{\lambda} - \frac{vt}{\lambda} \right),$$

and, if we also write ν for v/λ , Eqs. (8) and (9) become

$$E_y = A \sin 2\pi \left(\frac{x}{\lambda} - \nu t \right), \quad H_z = \sqrt{\frac{\epsilon}{\mu}} A \sin 2\pi \left(\frac{x}{\lambda} - \nu t \right). \quad (13a,b)$$

These equations represent sinusoidal waves of wave length λ and frequency ν , and of electric amplitude A . Their sinusoidal character can be brought out physically in two different ways. If we fix our attention on a particular point, *i.e.*, if we keep x constant, then at that point E_y and H_z are sine functions of the time t . On the other hand, if we take a "snapshot" of conditions at any instant, *i.e.*, if we keep t constant and vary x , E_y and H_z are sine functions of x .

Waves polarized in other planes, or traveling in other directions, can be obtained by writing down equations like (8) and (9) in terms of rotated axes and then changing variables back to the old x, y, z . In

this way can be obtained, for example,

$$E_z = f_1(x - vt), \quad H_y = -\sqrt{\frac{\epsilon}{\mu}} f_1(x - vt), \quad (14)$$

representing waves polarized with the electric vector in the xz plane, or

$$E_y = f_2(x + vt), \quad H_x = -\sqrt{\frac{\epsilon}{\mu}} f_2(x + vt), \quad (15)$$

representing waves of the original type but moving toward $-x$ instead of $+x$, and so on.

By superposing two wave trains traveling in the same direction but polarized in perpendicular directions, the most general type of plane waves traveling in a definite direction can be represented. At any point of such a train, as in the examples just given, the electric field E and the magnetic field H are perpendicular both to each other and to the direction of propagation; the three vectors E , H , and direction of propagation form a right-handed set (cf. Sec. 30 and Fig. 7). The numerical ratio of H to E at any point is $\sqrt{\epsilon/\mu}$ when E is measured in electrostatic units and H in oersteds. Thus, in free space, $H = E$.

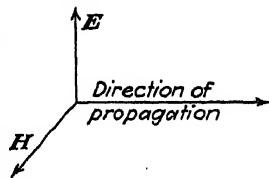


FIG. 7.—Relative directions in a plane wave.

It should be remarked, however, that in other types of mediums, such as crystals or conductors, the relationships of E and H are somewhat different from those just described.

It is well known that electromagnetic waves of the sort we have been describing can be produced in the laboratory; and it was supposed until recently that nothing had to be changed in the description except the wave length in order to have an accurate description of ordinary light. The discoveries of the present century, however, as we shall see in later chapters, have made necessary a considerable modification in our conception of the electromagnetic field, so that the picture just described, while sufficiently correct in all details for long electromagnetic waves, is not immediately applicable to light.

33. Velocity of Electromagnetic Waves.—The velocity of the waves that we have shown to represent a possible solution of the Maxwell equations is

$$v = \frac{c}{\sqrt{\epsilon\mu}}. \quad (16)$$

Thus in *free space*, where $\epsilon = \mu = 1$, the velocity of these waves equals c , the ratio of the electromagnetic to the electrostatic unit of

charge. When Maxwell announced this result in 1862, he had available Weber and Kohlrausch's value for c , the ratio of the units, published in 1857, and Fizeau's value of 1849 for v_0 , the velocity of light in a vacuum, *viz.*,

$$c = 3.11 \times 10^{10}, \quad v_0 = 3.14 \times 10^{10}.$$

This was a good agreement in those days. The best modern values are probably those furnished by the measurements of Rosa and Dorsey for c and the last measurement of v_0 made by Michelson, *viz.*,¹

$$c = 2.9979 \times 10^{10}, \quad v_0 = 2.99796 \times 10^{10}.$$

The agreement is within the experimental error, and the two quantities are universally regarded as exactly equal.

An interesting conclusion from this result is that the ratio of the units must have the dimensions of a velocity. It is easy to see more directly why this must be so. For a charge Q in *electrostatic* units exerts on another charge Q' , at a distance r , the Coulomb force

$$F_1 = \frac{QQ'}{r^2};$$

whereas a current element $i \, ds$, measured in *electromagnetic* units, exerts on another current element $i' \, ds'$ at a distance r , under suitable geometrical conditions, a force

$$F_2 = \frac{(i \, ds)(i' \, ds')}{r^2},$$

by the ordinary Biot formula for the magnetic field of a current element. The similarity of these formulas shows that Q and $i \, ds$ must have the same dimensions. But i represents charge per unit time; hence, $i \, ds$ has the dimensions of a charge measured in electromagnetic units times length divided by time, or times a velocity. It follows that the ratio of a charge Q measured in *electrostatic* units to a charge measured in *electromagnetic* units has the dimensions of a velocity. The number or *numeric* representing the measure of a quantity, however, has dimensions reciprocal to those of the corresponding *unit*; the numeric and the unit change in reciprocal ratios when the fundamental units of mass, length, and time are changed. Hence, it is the ratio of the *electromagnetic unit* to the *electrostatic unit* of charge that has the dimensions of a velocity.

In a material substance, on the other hand, the velocity of electromagnetic waves ought usually to be different from c . In any sub-

¹ Cf. BIRGE, *Rev. Mod. Phys.*, vol. 1, p. 1 (1929). But cf. also DUNNINGTON, *Rev. Mod. Phys.*, vol. 11, p. 65 (1939).

stance free from conductivity, $\mu = 1$ very nearly, so that (16) becomes

$$v = \frac{c}{\sqrt{\epsilon}}.$$

The refractive index, which is the ratio of the velocity in vacuum to the velocity in the substance, will be, therefore, for electromagnetic waves in such a substance,

$$n = \frac{c}{v} = \sqrt{\epsilon}, \quad (17)$$

or the refractive index should be equal to the square root of the dielectric constant. A comparison of the data now at hand shows that in many cases values of n for visible light agree closely with Eq. (17), but more commonly there are discrepancies, which may even be glaring. For example, for water, $\epsilon = 80$, $\sqrt{\epsilon} = 8.94$, but $n = 1.33$ for yellow light. Furthermore, Eq. (17) indicates no variation of n with wave length and thus leaves no room for the phenomenon of dispersion. The discrepancies tend to disappear, however, when sufficiently long waves are used. For example, for waves produced electrically with a wave length of 1.75 cm., water has an index of refraction of 7.82, whereas, for a wave length of 65 cm., the index is 8.88, in almost exact agreement with $\sqrt{\epsilon} = 8.94$.

The existence of dispersion indicates, as Maxwell himself suggested, that in dealing with material bodies certain factors have not been taken into account. Later work has shown that dispersion can be accounted for satisfactorily by assuming matter to contain electrical charges capable of vibrating elastically with definite natural periods. When the frequency of the disturbing electric field is very low, these charges undergo displacements proportional to the field strength very nearly to the same degree as they do in the steady field that is employed in measuring the dielectric constant; but at higher frequency their inertia modifies their motion, and their effect upon the waves is thereby altered. By assuming the motion of the particles to be damped in some way, and by including the effect of the free electrons in metallic conductors, it has been possible to give a good account of the optical properties of matter. To allow for all of these effects, the field equations as we have written them above have to be modified; furthermore, to obtain accurate results it is necessary to introduce quantum theory. We have not space, however, to pursue further here the subject of the electromagnetic theory of light.

34. Energy of Electromagnetic Waves.—The most important property of electromagnetic waves is that they convey energy.

According to Eq. (7) above, the energy is distributed in space at the rate of

$$w = \frac{\epsilon E^2 + \mu H^2}{8\pi} \text{ ergs per cm.}^3, \quad (18)$$

E and H being the electric and magnetic field strengths and ϵ and μ the dielectric constant and permeability, respectively. For plane waves traveling in a definite direction, this expression can be put into other forms by using the relations obtained from (9),

$$H = \sqrt{\frac{\epsilon}{\mu}} E, \quad v = \frac{c}{\sqrt{\epsilon\mu}}, \quad (19a,b)$$

v being the velocity of the waves. We note that for such waves $\mu H^2 = \epsilon E^2$, so that the amounts of electric and magnetic energies are equal, and we can also write

$$w = \frac{\epsilon E^2}{4\pi}. \quad (20)$$

This energy is carried along with the waves; and what is actually observable is usually not the energy density itself but the amount of energy that flows per second across unit area of a plane drawn perpendicular to the direction of propagation of the waves. This latter quantity is called the *intensity* of the train of waves; we shall denote it by I . The connection between I and w is easily found. Consider an elementary volume with faces of unit area and of thickness dx , placed perpendicular to the x -axis, the waves traveling as usual

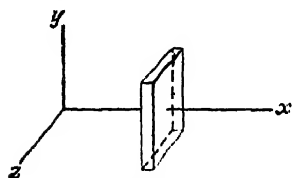


FIG. 8.

toward $+x$ (Fig. 8). The energy in this element is $w dx$, since its volume is dx . As the wave moves to the right with velocity v , in a time dt given by

$$dt = \frac{dx}{v},$$

all of this energy passes through the right-hand face of the element. The rate of flow of the energy across this face, which equals I because the face has unit area, is, therefore,

$$I = \frac{w dx}{dt}$$

or

$$I = vw. \quad (21a)$$

In free space, where $v = c$, this takes the form

$$I = cv. \quad (21b)$$

By means of (19a), (19b), and (20), Eq. (21a) can also be put into the form

$$I = \frac{c}{4\pi} EH; \quad (22)$$

or, if $\mu = 1$, we may write it, from (21a) and (20), inserting $v = c/n$ and $\epsilon = n^2$ from (17),

$$I = \frac{cn}{4\pi} E^2. \quad (23)$$

Here n is the refractive index, and I is the intensity of the waves or the energy that flows in a second across unit area. In free space, where $n = 1$, this becomes

$$I = \frac{c}{4\pi} E^2. \quad (23a)$$

Equation (22) is also easily obtained from Poynting's vector, which was described in a previous section.

These formulas for I hold, primarily, at each separate point along the wave train. They can be applied to the *average* intensity, however, merely by averaging the quantities that appear on the right in the equations, i.e., by inserting for w or EH or E^2 the time average of this quantity at any fixed point. In the case of sinusoidal waves, the average values are easily found in terms of the amplitude. Thus, if, as in Eq. (13a),

$$E = A \sin 2\pi \left(\frac{x}{\lambda} - \nu t \right),$$

then E^2 has the same average value during all successive cycles of period $T = 1/\nu$ seconds or during any integral number of cycles. Hence the average of E^2 is

$$\begin{aligned} \bar{E}^2 &= \frac{1}{T} \int_0^T E^2 dt = \frac{A^2}{T} \int_0^T \sin^2 2\pi \left(\frac{x}{\lambda} - \nu t \right) dt, \\ &= \frac{A^2}{T} \int_0^T \left[\frac{1}{2} - \frac{1}{2} \cos 4\pi \left(\frac{x}{\lambda} - \nu t \right) \right] dt \\ &= \frac{A^2}{T} \left[\frac{t}{2} - \frac{1}{8\pi\nu} \sin 4\pi \left(\frac{x}{\lambda} - \nu t \right) \right]_{t=0}^T = \frac{1}{2} A^2, \end{aligned}$$

the sine having the same value at times t and $t + T$ or $t + 1/\nu$. Hence

the average intensity of the waves is, from (23),

$$I = \frac{cn}{8\pi} A^2. \quad (24)$$

35. Momentum of Electromagnetic Waves. Radiation Pressure.

Besides energy, electromagnetic waves carry *momentum*. As stated in Sec. 31, the direction of this momentum is that of Poynting's vector, which, in the case of plane waves traveling in a definite direction, is readily seen to be the direction of propagation of the waves. The amount of momentum in unit volume, furthermore, is equal to $\epsilon\mu/c^2$ times the magnitude of Poynting's vector; and the latter represents the energy transferred across unit area per unit time, which we have denoted above by I . Hence, the density of momentum in the waves, symbol g , is

$$g = \frac{\epsilon\mu I}{c^2} = \frac{\epsilon\mu}{4\pi c} EH, \quad (25)$$

by (22). Or, by using (21a),

$$g = \frac{\epsilon\mu vw}{c^2} = \frac{w}{v}, \quad (25a)$$

by (19b); and in free space

$$g = \frac{w}{c}. \quad (25b)$$

The total amount of momentum associated with energy W carried by plane waves in free space is, therefore,

$$G = \frac{W}{c}. \quad (25c)$$

Thus waves moving in free space carry with them momentum equal to the energy divided by the speed of light (in c.g.s. units). This fact is of interest in dealing with atomic phenomena, especially the Compton effect. Another interesting observation is that the momentum is the same as would be present if the energy had a mass equal to w/c^2 grams and were moving at the speed c of the waves. This is a special case of the general interpretation of electromagnetic momentum described above.

Experimentally, however, the electromagnetic momentum in the waves will manifest itself only in reactions upon material bodies when the momentum in the field is changed. Thus, if a beam of light is deflected, or perhaps absorbed, by a material body, a force will be observed to act on the body, just as a baseball player, when he destroys the momentum of a moving baseball by stopping it with his hand, feels

a blow on the hand. The phenomena so produced are those of *radiation pressure*.

The magnitude of the pressure is easily calculated in any given case from the momentum brought up by the radiation. Since force equals rate of change of momentum, the pressure will equal the momentum delivered by the waves to unit area in a second. Since the waves move a distance v in a second, the momentum brought up will be that in a cylinder of unit cross section and length v , or vg c.g.s. units. If, then, the waves are entirely absorbed by the body, the radiation pressure on it will be, by (25) and (9),

$$p = vg = \frac{nI}{c} \text{ dynes per cm}^2, \quad (26)$$

n denoting the refractive index of the medium or c/v .

The radiation pressure can also be deduced by calculating from ordinary laws the forces that the electric and magnetic fields in the waves exert upon the charges and currents produced by them in material bodies.¹ A third method is to calculate it from the Maxwell stresses in the waves. These stresses are known to give correct values for the forces upon material bodies. The deduction from the momentum carried by the waves is the simplest, however, and we shall find that it is interesting in connection with the explanation of such atomic phenomena as the Compton effect.

That radiation falling on a body should exert a pressure upon it is by no means a concept peculiar to modern physics. Over 300 years ago, Kepler suggested that the curvature of comets' tails away from the sun might be due to radiation pressure—and modern astronomers believe that Kepler was right! Newton recognized that Kepler's suggestion was in harmony with the corpuscular theory of light (just as the pressure produced by a gas is due to the impact of its molecules). Experimentally, the subject was attacked as early as 1750, and at intervals thereafter, but with inconclusive or conflicting results, due for the most part to the disturbing action of the gas surrounding the illuminated surface. Finally, however, the disturbances due to the gas were definitely eliminated, and radiation pressure was discovered almost simultaneously by Lebedev in Europe (1900) and by E. F. Nichols and G. F. Hull in America (1903).² The pressure, in the experiments of Nichols and Hull, was only about 7×10^{-6} dyne per square centimeter (which is several times the pressure due to strong sunlight), but they were able to measure this small pressure with

¹ Cf. PLANCK, M., "Theory of Heat Radiation," Part II, Chap. I.

² Cf. NICHOLS and HULL, *Phys. Rev.*, vol. 17, pp. 29, 91 (1903).

sufficient precision to show that the observed values agreed within 1 per cent with the theoretical values.

36. Field of a Uniformly Moving Point Charge.—We have seen how to obtain a solution of Maxwell's equations representing plane electromagnetic waves. It is natural to consider next how such waves can be generated or stopped. This problem was not solved by Maxwell; a complete treatment of it requires elaborate mathematical procedures. The emission of radiation by a point charge is of some importance in dealing with the production of X-rays. Accordingly, the characteristics of such emission will be described presently; its characteristic features can be found in part by elementary methods. As a preliminary, however, it will be useful to consider first the field of a concentrated charge that is in *uniform* motion in a vacuum. The charge will be assumed to be concentrated at a point, but the results will also hold, at least approximately, if it is merely confined within a finite but negligibly small volume.

If the charge is stationary, its field is simply the familiar Coulomb field. When it moves, however, we can easily show that it must be surrounded by a magnetic field as well; and, in the special case of uniform motion along a straight line, we find that there exists a simple relation between the magnetic and electric fields.

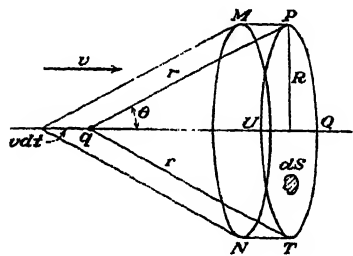


FIG. 9.

As the charge moves along at constant velocity, its field must move with it, so that relative to the charge the field seems to stand still. Any *magnetic* lines that may accompany it will have the form of circles having as their axis the line along which the charge is moving; for otherwise it is scarcely possible that the tubes of magnetic induction could be closed, as they must be. The field strength H will then have the same magnitude along any given line. The work done on a unit pole carried once around a line of radius R (e.g., $PUTQ$ in Fig. 9) will be, therefore,

$$W = 2\pi RH.$$

But, by the same circuital law that we employed above in deducing the first three of Maxwell's equations, this work is also equal to 4π times the current that links with the path, measured in electromagnetic units. We can calculate this total current entirely in terms of the displacement currents that are produced at each point in space as the electric field varies with time, due to the motion of the charge. To do

this, consider the plane circular area that is bounded by the line around which the pole was carried. Through any element dS of this area there flows a displacement current $j_{D1} dS$, j_{D1} being the component of the displacement-current density perpendicular to the area. The total current crossing the area inside the line is thus

$$i = \int j_{D1} dS,$$

integrated over the whole area. Now the value of j_D is given in Eq. (2), but in electrostatic units; to get it in electromagnetic units, we must divide by c . Hence, since here $\epsilon = 1$,

$$i = \frac{1}{4\pi c} \int \frac{dE_1}{dt} dS = \frac{1}{4\pi c} \frac{d}{dt} \int E_1 dS$$

where E_1 is the component of the electric field strength perpendicular to the area. But $\int E_1 dS = \varphi$, the total flux of electric induction through the part of the area bounded by the line. Hence,

$$i = \frac{1}{4\pi c} \frac{d\varphi}{dt};$$

and, equating $4\pi i$ to the expression found above for W , we obtain

$$2\pi R H = \frac{1}{c} \frac{d\varphi}{dt}. \quad (27)$$

Now the rate of change of φ is easily found, since it results merely from the convection of the field with the moving charge. Let the speed of the latter be v . We can imagine that the magnetic line which coincides momentarily with the circle PT in Fig. 9 is moving with the charge, the flux φ of electric induction through it remaining constant. At time t , let this line be in the position MN , and at time $t + dt$ let it be in the position PT . Then $MP = v dt$. Now, at time $t + dt$, all of φ passes through PT ; whereas, at time t , part of it escaped between the two circles. Thus in dt the flux through PT increases by the amount of flux which at time t passed outward through the cylindrical surface $MPTN$. Since the area of this surface is

$$2\pi R \times (MP) = 2\pi R v dt,$$

we have, therefore, for the increase in the flux or $d\varphi$, $2\pi R v E_{\perp} dt$, where E_{\perp} stands for the component of the electric field perpendicular to the direction of motion. Hence

$$\frac{d\varphi}{dt} = 2\pi R v E_{\perp}.$$

From this equation and (27), we have, then,

$$H = \frac{v}{c} E_{\perp}. \quad (28)$$

This equation gives us the magnetic field H in oersteds at any point in terms of the electric field E measured in electrostatic units.

It remains to find the electric field E . Now if the motion is *very slow* as compared with the speed of light (v/c very small), the electric field will be sensibly the same as that of a stationary charge. Hence, at a point distant r from the charge in a direction making an angle θ with its direction of motion (Fig. 9),

$$E_{\perp} = \frac{q}{r^2} \sin \theta, \quad (29)$$

q being the charge in electrostatic units. Thus, at such a point, by (28),

$$H = \frac{qv}{cr^2} \sin \theta. \quad (30a)$$

Or, if q' is the charge in electromagnetic units, H being as before in oersteds,

$$H = \frac{q'v}{r^2} \sin \theta. \quad (30b)$$

From this formula we see that the magnetic field around a slowly moving point charge is the same as the field that is due, according to the familiar Biot (or "Ampère's") formula, to a current element of magnitude $i ds = qv/c = q'v$.

If the speed v of the charge is not small as compared with the velocity of light, we have to take account of the modification produced in the electric field by the varying magnetic flux (varying, that is, at points fixed relative to the frame of reference used in defining the speed of the particle). Unfortunately, this modification cannot be found by elementary methods. It is most conveniently found by a relativity transformation, and this will be done in a later chapter on "Relativity."

It turns out that the electric lines remain straight at all speeds, pointing radially away from the charge in its instantaneous position, but the magnitude of the field strength, as shown in Sec. 70, is

$$E = \frac{\gamma}{[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}}} \frac{q}{r^2}, \quad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (31a,b)$$

Then, by (28) and (29),

$$H = \frac{\gamma}{[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}}} \frac{qv}{cr^2} \sin \theta. \quad (32)$$

This formula shows that, as v increases, the whole field tends to become compressed toward an "equatorial" plane drawn through the charge and perpendicular to its motion. On the line of motion, for example, where $\theta = 0$ and $[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}} = \gamma^3$, E is less in the ratio $1/\gamma^2$ than it is at the same distance from a stationary charge; whereas, in the equatorial plane, where $\theta = 90^\circ$ and $[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}} = 1$, it is greater in the ratio γ . As $v \rightarrow c$, the field tends to be concentrated entirely in the equatorial plane.

37. Radiation Field of an Accelerated Point Charge.—As we have seen, the field around a uniformly moving point charge is simply carried along with the charge, so that the energy in such a field remains in the neighborhood of the charge. Thus, from uniformly moving charges, there is no radiation of energy, for "radiation" implies that energy passes off to infinity.

The situation is quite different in regard to *accelerated* charges. A complete treatment of this more general case is beyond the scope of this book; and here relativity lends no help. It happens, however, that only one conclusion need be cited from the advanced theory. It can be shown that the field due to any element of charge may be supposed to be emitted continually from that element and to spread out in all directions away from it with the speed of light.¹ This being assumed, the rest of the deduction requires only elementary methods.

Consider the special case of a point charge q , which for some time has been at rest at a point O in free space (Fig. 10) and now undergoes a uniform acceleration a during a very short time τ , after which it continues to move uniformly with velocity

$$v = a\tau.$$

At the end of the period of acceleration, it will have moved to a slightly different point P_1 ; and, after an additional time t , much greater

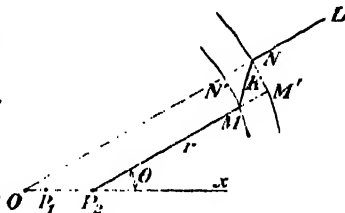


FIG. 10. — Diagram illustrating the emission of a pulse of radiation.

¹ Cf. ABRAHAM, M., and R. BECKER, "Electricity and Magnetism," p. 220 ("Electrodynamic Potentials"); W. R. SMYTHE, "Static and Dynamic Electricity," Sec. 13.25 ("Retarded Potentials").

than τ , it will have got to P_2 , where

$$P_1P_2 = vt.$$

Let us find the field at the end of the additional time t , when the charge is at P_2 .

Let ONL represent any line of force drawn from the charge while stationary at O ; let N lie on a sphere of radius $c(\tau + t)$ about O as center, c being the speed of light. The field emitted at the beginning of the period τ has just had time to reach the surface of this sphere. The field *outside* must have been emitted earlier, and, therefore, according to the conclusion just cited from the exact theory, it must be the original field associated with the stationary charge at O , not having received "news" of the acceleration of the charge. Similarly, *inside* a sphere of radius ct with center at P_1 , we have the field of the charge moving uniformly with speed v . Between these two spheres is a shell within which the transition from the one type of field to the other occurs, the line of force running obliquely here from M to N . If we allow time to flow on, this transition shell moves outward with the speed of light. We shall show that it contains a radiation component of the field that carries a certain amount of energy off to infinity.

The part MN of the force line is straight, at least in the limit as $\tau \rightarrow 0$, because the acceleration is uniform.¹ Let us resolve the electric field at any point on MN into a radial component E_r , in the direction of P_2M , and a perpendicular component E_t . Then, at any point on MN ,

$$\frac{E_t}{E_r} = \frac{M'N}{MM'} = \frac{MN'}{MM'}.$$

But P_2M is (in the limit as $\tau \rightarrow 0$) parallel to ON' , since, according to (31a) above, the change in the electric field due to the motion is of the second order in v , and since P_2M and ON' are different positions of the same force line (a line can be identified by its lying on a cone about Ox which encloses a given fraction of the electric flux of induction). Hence, if we neglect the distance OP_1 , which is of the second order in v ,

$$MN' = (OP_2) \sin \theta = vt \sin \theta = at \sin \theta.$$

¹ That MN must be straight (in the limit) is easily seen, after reading the whole deduction, from the following consideration. The field at any point K on MN was emitted at a certain time τ' after the charge left O . Had the acceleration ceased at this instant, the point M , as defined in the text, would have been at K and the distance of M or K from ON , according to the expression found in the text for MN' , would have been $a\tau' \sin \theta$. Thus the distance of K from ON is proportional to τ' . But so is KN , to the first order. Hence the line MN is straight.

Furthermore, MM' equals the difference in radii of the spheres if OP_1 is again neglected, whence,

$$MM' = ct.$$

Therefore,

$$\frac{E_t}{E_r} = -\frac{at}{c} \sin \theta;$$

a minus sign has been inserted here because E_t has a component opposite to the acceleration a . But E_r must have the same value in the shell as outside in order to carry the flux of induction through the shell and so satisfy Gauss' theorem. Hence,

$$E_r = \frac{q}{r^2},$$

r being the radius of the shell. Thus,

$$E_t = -\frac{q}{r^2} \frac{at}{c} \sin \theta,$$

or, since $r = ct$, so that $t = r/c$,

$$E_t = -\frac{qa}{c^2 r} \sin \theta. \quad (33)$$

Thus, within the transition shell there is a tangential electric component E_t , which falls off, as r increases, only as the *first power* of r . This component lies in a plane drawn through the charge and containing the direction of its acceleration. As time goes on and $r \rightarrow \infty$, E_r will become entirely negligible in comparison with E_t . Thus, ultimately, we have a shell or pulse of field, traveling outward with the speed of light, in which the electric vector is *transverse*, just as in plane waves. In fact, any portion of the pulse ultimately approximates a plane wave. Nothing has been said about the magnetic field, which is not so easily treated; but it can scarcely be doubted that there must also exist in the pulse a transverse magnetic field, equal in strength to the electric field but perpendicular to it, just as we found to exist in plane waves, and this conclusion is confirmed by the complete theory. The electromagnetic field that thus travels away to infinity from an accelerated charge is often called its *radiation field*.

By means of a relativity transformation, these results can readily be generalized so as to represent the radiation from a charge that is already in motion. It turns out that the radiation field is always numerically proportional to the *acceleration* of the charge, and in this field E , measured in electrostatic units, and H , measured in oersteds,

are equal; furthermore, E , H , and the radius from the point of emission form a right-handed set of mutually perpendicular vectors (*cf.* Sec. 30). The formula for E is somewhat complicated, in general, so we shall write it down only for the simple case in which the acceleration and the velocity lie along the same line, as in *linear motion*; in this case¹

$$E = -\frac{1}{(1 - \beta \cos \theta)^2} \frac{qa}{c^2 r} \sin \theta, \quad \beta = \frac{v}{c}, \quad (34a, b)$$

v being the instantaneous speed and a the instantaneous acceleration of the charge. We observe that the radiation field vanishes, in this case, at points on the line of motion, and that E lies in a plane drawn through the line of motion, just as for a slowly moving charge. The only effect of the motion, in fact, is represented by the first factor on the right in (34a), which serves to strengthen the field in forward directions and to weaken it toward the rear of the moving charge.

38. Energy Radiated by an Accelerated Point Charge.—The most important feature introduced by the acceleration of a charge is the resulting radiation of energy. We can readily calculate its magnitude for a charge moving along a line, from the formula just given.

Let a charge, of magnitude q in electrostatic units, move with acceleration a , constant or variable, along a line Ox (Fig. 11). During an interval of time dt , let the charge move at speed v from P to P' , so that $PP' = v dt$. Then

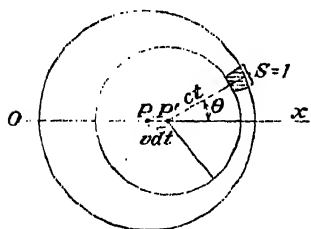


FIG. 11.—Diagram illustrating a pulse of radiation.

the radiation field emitted during this interval will be found, at an instant t seconds after the time at which the charge passed P' , between two spheres, one of radius ct about P' as center, the other of radius $c(t + dt)$ about P as center. Let the time t be large enough so that in this region there is only the radiation field, in which $H = E$. Then there is in this field an energy density of magnitude (vacuum being assumed)

$$w = \frac{E^2 + H^2}{8\pi} = \frac{E^2}{4\pi}. \quad (35)$$

The distance between the two spheres, which would be $c dt$ if they had a common center, is changed by their relative displacement to the extent of the projection of PP' on the radius; thus, in a direction making an angle θ with the motion, the distance between the spheres is $c dt - v dt \cos \theta$ or $(c - v \cos \theta) dt$. Hence, between each pair of

¹ *Cf.* SMYTHE, *op. cit.*, Sec. 14.12.

opposing unit areas on the spheres, there lies a volume equal to

$$(c - v \cos \theta) dt$$

times unity; and this volume contains energy of amount

$$dW_1 = w (c - v \cos \theta) dt \text{ ergs.}$$

Inserting the value of w from (35) and of E from (34a), we find

$$dW_1 = \frac{1}{4\pi} (c - v \cos \theta) \frac{q^2 a^2}{c^4 r^2} \frac{\sin^2 \theta}{(1 - \beta \cos \theta)^5} dt$$

or, dividing out c and using $\beta = v/c$ [Eq. (34b)],

$$dW_1 = \frac{1}{4\pi} \frac{q^2 a^2}{c^3 r^2} \frac{\sin^2 \theta}{(1 - \beta \cos \theta)^5} dt. \quad (36a)$$

If the velocity v is much below that of light, so that β is negligible, we can write this in the form

$$dW_1 = \frac{1}{4\pi} \frac{q^2 a^2}{c^3 r^2} \sin^2 \theta dt. \quad (36b)$$

This energy dW_1 , lying between two unit areas on the spheres, eventually passes outward across a fixed unit area held so as to coincide momentarily with that on the outer sphere. Thus, dW_1 represents the radiant energy that passes outward in time dt across a unit area located at a distance r and in the direction specified by θ from the accelerated charge.

The formula for dW_1 is applied to particular problems in different ways.

(a) *Instantaneous Rate of Radiation of Energy.*—The total rate at which energy is passing outward or is being radiated can be found by integrating over the outer sphere, whose radius is sensibly equal to r or to ct . We shall do this for the case of slow motion only. As an element of area on the sphere, take a ring of width $r d\theta$ and radius $r \sin \theta$, and hence of area $2\pi r^2 \sin \theta d\theta$, using as the axis of this ring the line from which θ is measured. The value of dW_1 is uniform over the ring, hence the energy that passes outward across it in time dt is

$$2\pi r^2 \sin \theta dW_1 d\theta = \frac{q^2 a^2}{2c^3} \sin^3 \theta d\theta dt$$

by Eq. (36b). The total energy that passes outward over the whole sphere is, therefore, by integration,

$$\frac{q^2 a^2}{2c^3} dt \int_0^\pi \sin^3 \theta d\theta.$$

Now

$$\int_0^\pi \sin^3 \theta d\theta = \int_0^\pi (1 - \cos^2 \theta) \sin \theta d\theta = \left[-\cos \theta + \frac{1}{3} \cos^3 \theta \right]_0^\pi = \frac{4}{3} \quad (37)$$

Hence, dividing by dt , we have for the energy passing outward over the sphere per second

$$\Omega = \frac{2}{3} \frac{q^2 a^2}{c^3}. \quad (38)$$

This is the rate at which the charge q is radiating energy in all directions at an instant when its acceleration has the value a .

(b) *Stopping of a Particle.*—Suppose a particle carrying a charge q and moving initially with speed v_0 is brought to rest without change in its direction of motion. This case is of interest in connection with the production of X-rays. An adequate theory of X-rays can be given only in terms of quantum theory, but the results of classical theory furnish useful analogies.

Consider the radiation emitted by the charge during the stopping process, in a direction making an angle θ with the direction of motion. At a large distance r , the total energy W_1 crossing unit area of a plane perpendicular to the radius from the particle can be got by integrating dW_1 , as given in (36a) or (36b), with respect to the time. The result of the integration will depend upon the manner in which the acceleration a varies with the time; for simplicity, let the acceleration be uniform. Then, if v_0 is small compared with the speed of light c , since for the process in question

$$\int dt = \int_{v_0}^0 \frac{dv}{a} = -\frac{v_0}{a},$$

we have from (36b)

$$W_1 = \int a W_1 = \frac{1}{4\pi} \frac{q^2 v_0 |a|}{c^3 r^2} \sin^2 \theta, \quad (39a)$$

$|a|$ being the absolute value of the acceleration and equal to $-a$. If, on the other hand, v_0/c is not small, we have only to note that $dt = dv/a = c d\beta/a$, so that, writing $\beta_0 = v_0/c$, we have during the stopping process

$$\int \frac{dt}{(1 - \beta \cos \theta)^5} = \frac{c}{a} \int_{\beta_0}^0 \frac{d\beta}{(1 - \beta \cos \theta)^5} = \frac{c}{4a \cos \theta} \left[1 - \frac{1}{(1 - \beta_0 \cos \theta)^4} \right].$$

By noting that

$$\begin{aligned} 1 - (1 - \beta_0 \cos \theta)^4 &= [1 - (1 - \beta_0 \cos \theta)^2][1 + (1 - \beta_0 \cos \theta)^2] \\ &= (2\beta_0 \cos \theta - \beta_0^2 \cos^2 \theta)(2 - 2\beta_0 \cos \theta + \beta_0^2 \cos^2 \theta), \end{aligned}$$

the last expression can be got into a more useful form. The result of integrating (36a) can then be written

$$W_1 = \frac{1}{4\pi} \frac{q^2 v_0 |a|}{c^3 r^2} \sin^2 \theta \frac{(1 - \frac{1}{2} \beta_0 \cos \theta)(1 - \beta_0 \cos \theta + \frac{1}{2} \beta_0^2 \cos^2 \theta)}{(1 - \beta_0 \cos \theta)^4}. \quad (39b)$$

The *total* radiation emitted during the stopping process can then also be found by integrating W_1 in turn over a sphere of radius r . We shall do this, again, for the case of slow motion only. As an element of area on the sphere, take, as before, a ring of width $r d\theta$ and radius $r \sin \theta$, and hence of area $2\pi r^2 \sin \theta d\theta$, its axis being the line from which θ is measured. W_1 is uniform over this ring. Hence the total energy emitted is, from (39a),

$$W = 2\pi \int_0^\pi W_1 r^2 \sin \theta d\theta = \frac{q^2 v_0 |a|}{2c^3} \int_0^\pi \sin^3 \theta d\theta,$$

or, by (37),

$$W = \frac{2}{3} \frac{q^2 v_0 |a|}{c^3}. \quad (39c)$$

We note that, the more rapidly the charge is brought to rest the larger is the amount of energy radiated.

(c) *Damped Harmonic Oscillator*.—A case of interest in connection with the emission of radiation from atoms is that of an electron or other particle bound elastically to an equilibrium position. If the particle is set into vibration by some disturbance, it will gradually lose its energy by radiation because of the acceleration experienced at various parts of its path and will eventually come to rest. It should be remarked again that correct results for atomic phenomena can be obtained only from quantum theory; but the analogous phenomenon in classical theory is interesting.

Suppose for simplicity that the motion is one-dimensional. Then, if the damping by radiation is small, the displacement x of the particle from its position of equilibrium will be given approximately by the usual equation of simple harmonic motion,

$$x = A \sin 2\pi\nu(t + \delta),$$

and its energy of vibration will be

$$W = \frac{1}{2} m (v_{\max})^2 = 2\pi^2 m \nu^2 A^2, \quad (40)$$

v_{\max} , or $2\pi\nu A$ being its maximum velocity and m its mass. The rate of loss of energy by the oscillator can be found from (36b) if we suppose the velocity to be always small compared with that of light. We integrate over a large sphere as was done in deducing (38). This gives,

for the energy radiated during time dt ,

$$2\pi \int_0^\pi dW_1 r^2 \sin \theta d\theta = \frac{q^2 a^2}{2c^3} dt \int_0^\pi \sin^2 \theta d\theta = \frac{2}{3} \frac{q^2 a^2}{c^3} dt$$

by (37), q being the charge on the particle. Now the acceleration is

$$a = \frac{d^2 x}{dt^2} = -4\pi^2 \nu^2 A \sin 2\pi \nu (t + \delta).$$

The energy lost during a single period of vibration, lasting $1/\nu$ seconds, will be, therefore,

$$\begin{aligned} \int \frac{2}{3} \frac{q^2 a^2}{c^3} dt &= \frac{32}{3} \frac{\pi^4 \nu^4 q^2 A^2}{c^3} \int_{t_1}^{t_1+1/\nu} \sin^2 2\pi \nu (t + \delta) dt \\ &= \frac{16}{3} \frac{\pi^4 \nu^3 q^2 A^2}{c^3}, \quad (40a) \end{aligned}$$

since

$$\int_{t_1}^{t_1+1/\nu} \sin^2 2\pi \nu (t + \delta) dt = \int_{t_1}^{t_1+1/\nu} \left[\frac{1}{2} - \frac{1}{2} \cos 4\pi \nu (t + \delta) \right] dt = \frac{1}{2\nu}.$$

(Here t_1 is any fixed time that may be selected.) The expressions obtained in Eq. (40a), divided by the period $1/\nu$, represent the average loss of energy per second. Hence, we can write for the energy of the oscillator, if the loss in a single period is small, the approximate differential equation

$$\frac{dW}{dt} = -\frac{16}{3} \frac{\pi^4 \nu^4 q^2 A^2}{c^3} = -\frac{8}{3} \frac{\pi^2 \nu^2 q^2}{mc^3} W, \quad (40b)$$

by Eq. (40). The solution of this equation is, if W_0 is the value of W at $t = 0$,

$$W = W_0 e^{-\alpha t}, \quad \alpha = \frac{8}{3} \frac{\pi^2 \nu^2 q^2}{mc^3}. \quad (41a, b)$$

The energy and amplitude of the vibrating charge thus decrease exponentially at a rate proportional to ν^2 .

In subsequent chapters, we shall find a number of uses for the formulas that we have been obtaining. The most important single conclusion from our work, however, is that *according to classical electromagnetic theory an accelerated charge must necessarily radiate energy.*

39. Electromagnetic Mass.—It was pointed out by J. J. Thomson in 1881 that a body should appear to have a larger mass when charged than it has when uncharged. The effect is essentially one of self-induction; just as the process of starting a current in a wire calls into

existence a momentary electromotive force opposing the current, so, similarly, setting a charge into motion produces, associated with the changing magnetic field around it, an electric field that acts on the charge so as to oppose its acceleration. The effect is thus the same as if the mass were increased.

Nowadays, however, we prefer to view this phenomenon in terms of the electromagnetic momentum in the field. The momentum can be calculated if the electric and magnetic fields around the body are known. The calculation is easily made for the case of a uniformly charged sphere moving *slowly* (as compared with the speed of light).

At a point distant r from the center of the sphere, on a radius making an angle θ with the direction of motion, we have a radial electric field strength

$$E = \frac{q}{r^2},$$

q being the total charge on the surface of the sphere in electrostatic units, and also a magnetic field whose magnitude, according to Eq. (30a) in Sec. 36, is

$$H = \frac{qv}{cr^2} \sin \theta.$$

Thus Poynting's vector, described in Sec. 31, will lie in a plane through the line of motion and will be perpendicular to the radius, as shown by the arrow marked P in Fig. 12; its magnitude, since E and H are perpendicular, will be

$$P = \frac{c}{4\pi} EH = \frac{q^2 v}{4\pi r^4} \sin \theta.$$

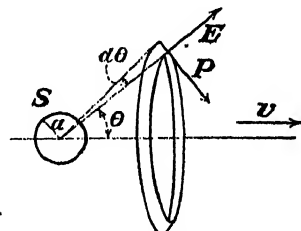


FIG. 12. Diagram illustrating the calculation of the electromagnetic momentum around a moving charged sphere S .

According to the rule stated in Sec. 31, therefore, there will be electromagnetic momentum in the region surrounding the sphere to the extent of P/c^2 per unit volume, its direction being that of Poynting's vector. Inside the sphere there is no field and hence no momentum.

Because of the symmetry, however, the component of this momentum perpendicular to the motion will sum up to zero. Thus the resultant momentum in the whole field is parallel to the motion, and to find it we need sum up components in this direction only. The component of electromagnetic momentum parallel to the motion is, per unit volume,

$$\frac{P}{c^2} \sin \theta = \frac{q^2 v}{4\pi c^2 r^4} \sin^2 \theta.$$

As an element of volume we may take a ring whose axis is the line of motion of the charge, with a cross section $r d\theta dr$ and a perimeter $2\pi r \sin \theta$ ($r \sin \theta$ being its radius), so that its volume is $2\pi r^2 \sin \theta d\theta dr$. The component of the momentum has the same value at all points on this ring. Then the total electromagnetic momentum in the field will be

$$\begin{aligned} M &= \int \int \frac{q^2 v}{4\pi c^2 r^4} \sin^2 \theta \cdot 2\pi r^2 \sin \theta d\theta dr \\ &= \frac{q^2 v}{2c^2} \int_a^\infty \frac{dr}{r^2} \int_0^\pi \sin^3 \theta d\theta, \end{aligned}$$

a being the radius of the sphere; or, from (37), and since $\int_a^\infty \frac{dr}{r^2} = \frac{1}{a}$,

$$M = \frac{2}{3} \frac{q^2 v}{ac^2}.$$

This is the same amount of momentum that the sphere would possess, according to the ordinary formula for momentum, if it had a mass m_q of magnitude

$$m_q = \frac{2}{3} \frac{q^2}{ac^2}; \quad (42a)$$

for then $M = m_q v$. The sphere will behave dynamically as if it possessed a mass $m = m_0 + m_q$, where m_0 is its ordinary mass.

The formula just obtained for m_q is correct, strictly speaking, only for indefinitely slow motion. In case the sphere becomes contracted in the direction of motion in the ratio $(1 - v^2/c^2)^{1/2}$, in accordance with the requirement of Relativity (cf. Sec. 64), it can be shown that the electromagnetic mass is

$$m_q = \frac{2}{3} \frac{q^2}{ac^2} \frac{1}{\sqrt{1 - v^2/c^2}}. \quad (42b)$$

The calculation of m_q for slow motion is often made by calculating the energy in the magnetic field around the sphere and equating this energy to $\frac{1}{2} m_q v^2$. This method would be reliable if it could be seen in some way that there is no change in the energy of the *electric* field arising from its distortion as represented by Eq. (31a) in Sec. 36. Actually there is no change of order v^2 in this part of the energy, so that a correct result is obtained for m_q ; but this comes about only because the decrease in electrical energy in polar regions, which is of the same order as the total magnetic energy, is offset by a nearly equal increase of energy in equatorial regions.

CHAPTER. III

THE PHOTOELECTRIC AND THERMIONIC EFFECTS

An approach to the outstanding problem of the first quarter of the present century, *viz.*, the nature of radiant energy, is most directly made by a study of the photoelectric effect. The term "photoelectric" might be applied with reason to a wide variety of phenomena involving the interaction between light and electricity, such as the change in the resistance of selenium under illumination or the rotation of the plane of polarization of a beam of light when passing through a medium placed in a strong electric field. It is commonly restricted, however, to the discharge of negative electricity from bodies when illuminated by light of appropriate wave length. In this chapter, we shall give a brief discussion of this effect, emphasizing those features which proved especially difficult of explanation in terms of the wave theory of light. The discussion of the photoelectric effect naturally involves a consideration of the experiments that led to one of the most striking advances of the modern period, the discovery of the electron. Some of these experiments will be described. Incidentally, we shall include a brief discussion of a closely related phenomenon, that of thermionic emission.

40. Discovery of the Photoelectric Effect. -

The photoelectric effect was discovered by Heinrich Hertz,¹ quite accidentally, in 1887. While carrying on experiments in which he demonstrated the existence of electromagnetic waves, he observed that the spark between the terminals (*S*, Fig. 13) of his detecting circuit passed more rapidly when those terminals were illuminated by the light from the primary spark *P* than when an opaque screen, or even a piece of plate glass, was interposed between *P* and *S*. Because the primary spark was known to emit ultraviolet light, which would be filtered out by the glass, Hertz concluded that the effect was due to ultraviolet light falling on the secondary gap—a conclusion which he confirmed by screening *S* from *P* and allowing light from another source to fall on *S*. Further, he found that the light must fall on the terminals themselves, which,

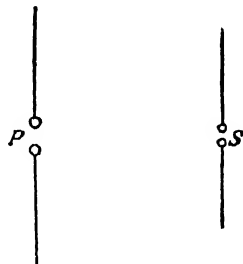


FIG. 13.

¹ *Ann. d. Physik*, vol. 31, p. 983 (1887).

for the effect to be observed, must be "smooth and clean." Hertz announced this discovery in a paper entitled "An Effect of Ultraviolet Light upon the Electric Discharge," and then returned to his investigations of electromagnetic waves.

Hertz's discovery at once attracted numerous investigators. Hallwachs¹ found that a freshly polished zinc plate, insulated and connected to an electroscope as an indicator, when charged *negatively* and illuminated by ultraviolet light, would lose its charge, but that there was no effect if the charge was *positive*. He concluded that in some mysterious manner, when the plate was negatively charged and illuminated, negatively electrified particles were emitted from it. He even observed, by using an electrometer instead of an electroscope, that a *neutral* insulated plate, when illuminated, would acquire a small *positive* potential, i.e., would lose a *negative* charge. Not only did light *permit the escape* of negative electricity from a negatively electrified plate, it even *caused the expulsion of negative electricity from a neutral plate*.

Stoletow² devised an arrangement, shown diagrammatically in

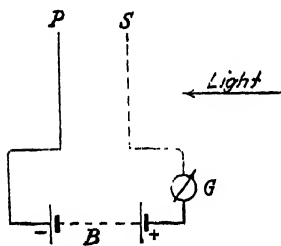


FIG. 14.

Fig. 14, for producing a continuous *photoelectric current*. P is a photoelectrically sensitive plate, say a polished zinc plate, connected to the negative terminal of a battery B of several cells. S is a wire grating or gauze connected to the positive terminal of the battery through a very sensitive galvanometer or electrometer G. When ultraviolet light falls upon P, a continuous current is observed in G, indicating that a negative

charge is flowing from P to S. No current flows if the battery is reversed.

Elster and Geitel³ showed that there is a close relation between the contact potential series of metals and the photoelectric effect: the more electropositive the metal, the longer the wave length to which it would respond photoelectrically. The alkali metals, sodium, potassium, and rubidium, were found to be sensitive to light of the visible spectrum. It was early observed that the photoelectric effect took place even in the highest attainable vacuum. For example, if the plates P and S, in Fig. 14, were enclosed in a glass bulb fitted with a quartz window for admitting ultraviolet light, the photoelectric

¹ *Ann. d. Physik*, vol. 33, p. 301 (January, 1888).

² *J. de. Physique*, vol. 9, p. 486 (1890).

³ *Ann. d. Physik*, vol. 38, pp. 40, 497 (1889).

current was observed to flow even if the bulb was evacuated as thoroughly as possible. This showed that the gas surrounding the plate played no essential role.

41. A Problem.—An important question then arose. What is the mechanism by which negative electricity is transferred from the cathode to the anode? That the charge is carried by negatively electrified particles was clearly indicated by experiments of Elster and Geitel, who showed that a transverse magnetic field diminishes the photoelectric current if the phenomenon takes place in a vacuum. But what are the particles?

Negative answers to this question were readily obtained. The fact that the effect persisted even to the highest attainable vacuum and was quite independent of the "degree" of the vacuum after a certain low pressure had been reached seemed to indicate that the gas molecules themselves in the region between *P* and *S* could not be acting as carriers of the charge. The suggestion was made that, perhaps, under the influence of light, negatively charged particles of the cathode became detached and moved to the anode. This suggestion was rendered untenable by an experiment by P. Lenard,¹ in which a clean platinum wire acted as anode and a sodium amalgam as cathode, both being in an atmosphere of hydrogen. The photoelectric current was allowed to flow until about 3×10^{-6} coulomb had passed through the circuit. If the carriers of the charge were atoms of sodium, each atom could hardly be expected to carry a larger charge than it carries in electrolysis. Taking the electrochemical equivalent of sodium as 0.00024 gram per coulomb, there should have been deposited on the platinum wire at least 0.7×10^{-6} milligram of sodium, a quantity sufficient to be detectable by the well-known flame test. On removing the wire from the bulb, however, no trace of sodium could be detected. Evidently, particles of the cathode did *not* act as carriers of the photoelectric current.

If the photoelectric current is carried neither by molecules of the gas surrounding the cathode nor by molecules of the cathode itself, what are the carriers? The answer to this question came as a result of the convergence of a number of different lines of evidence, which finally culminated in the discovery of the electron by J. J. Thomson. Before continuing the discussion of the photoelectric effect we shall consider some of these other developments.²

¹ *Ann. d. Physik*, vol. 2, p. 359 (1900).

² Collectively these developments illustrate many characteristics of the growth of modern physics:

1. A number of seemingly unrelated lines of research frequently converge to

42. Electricity in Matter.—A certain atomicity of electricity was suggested long ago by the laws of electrolysis. Faraday found that when the same quantity of electricity deposits univalent ions of different kinds, the amounts deposited, as measured by weight, are proportional to the *ionic weights* in the chemical sense. By the “ionic weight” is meant the ordinary atomic weight (oxygen = 16) if the ion consists of a single atom, or the sum of the atomic weights of the atoms composing a compound ion. Suppose an amount F of electricity deposits a gram-ion of a certain kind of univalent ion, *i.e.*, a quantity the mass of which in grams is equal to the ionic weight. If the ion is monatomic, the mass deposited will be a gram-atom, or a number of grams equal to the atomic weight. Then the same quantity of electricity would suffice to deposit a gram-ion of any other kind of univalent ion. This quantity of electricity is an important fundamental constant in chemistry and physics and is called the *faraday*. The best modern value of it is¹

$$F = 96,489 \text{ coulombs.}$$

Now the number of ions in a gram-ion, or of atoms in a gram-atom (or of molecules in a gram-molecule) is in all cases the same; for, if equal numbers be taken of different kinds of ions, the masses will be proportional to the ionic weights of the ions. This number, which we shall denote by N_0 , is an important physical constant. It follows that all univalent ions carry the same numerical charge, equal to F/N_0 . The charge on a univalent ion thus constitutes a natural unit of charge,

provide an explanation of, or a theory for, a group of phenomena not hitherto understood.

2. The explanation or theory thus evolved is then found to bear directly on other branches of physics and often on other sciences.

3. Thus, the methods of physics are both (1) synthetic and (2) analytic—a fortunate circumstance, which makes it possible for the physicist to comprehend physics as a whole in spite of the vast increase, particularly in recent years, of factual knowledge.

4. These discoveries, of both fact and theory, are the *sine qua non* of applied physics and of much of industry—witness, as a single example, the wide use of the various kinds of photo- and thermionic tubes.

5. This application of science to industry very frequently reacts to provide the research man with improved tools. Electron tubes, manufactured primarily for industrial purposes, are a boon to the physicist in his research laboratory.

6. In all these developments, scientists of all nations “collaborate,” the more effectively because of the very fact that they are not formally organized, *let alone directed*, but are free as individuals to follow their respective interests.

An understanding of these generalizations is essential to an understanding of modern physics.

¹ BRIDGE, *Rev. Modern Physics*, vol. 1, p. 1 (1929).

usually denoted by e . Multivalent ions carry multiples of e , divalent ions carry charges of $2e$, and so on. For this natural unit of charge Dr. G. Johnstone Stoney proposed in 1891 the name "electron."

If the number N_0 were known, we could at once calculate e from the equation

$$e = \frac{F}{N_0}.$$

In Stoney's time, however, it was known only, from kinetic theory, that N_0 was very large, perhaps of the order of 10^{23} , so that e was of the order of 10^{-19} electromagnetic unit. We shall see later that, as a matter of fact, N_0 is best determined by first measuring e and then calculating N_0 , as $N_0 = F/e$.

An independent line of thought which likewise suggested the presence of definite electrical charges in matter was the treatment of dispersion in developing the electromagnetic theory of light. L. Lorenz suggested in 1880¹ that refractive mediums may contain small charged particles which can vibrate with a natural period ν_0 about a fixed equilibrium position. By assuming one or more sets of such particles, with different natural frequencies, it was possible to account completely for the phenomena of dispersion. But the data on dispersion could not be made to furnish any clue as to the magnitude of the charges on these particles, as to their mass, or even as to the sign of their charge. The first evidence on this point from optical phenomena was furnished by a new effect discovered by Zeeman in 1897.

43. The Zeeman Effect.—In 1862, Faraday, looking for a possible effect of a magnetic field upon a light source, placed a sodium flame between the poles of a strong electromagnet and examined the D lines by a spectroscope. He was unable to detect any change in the appearance of the lines.

Faraday's failure to observe the effect that he expected was due to the inadequate resolving power of his apparatus. For, in 1896, Zeeman,² repeating Faraday's experiment with the improved technique then available, discovered that spectral lines are split up into components when the source emitting the lines is placed in a very strong magnetic field; furthermore, following a suggestion by H. A. Lorentz, he found that these components are polarized. The simplest case is shown in Fig. 15, where a represents a line before the magnetic field is turned on. If the field is turned on and the line viewed *at right angles* (R) to the direction of the field, the line is seen to be triple with components l_R , a' , and r_R . The central line a' has the same wave

¹ LORENZ, *Ann. d. Physik*, vol. 11, p. 70 (1880).

² *Phil. Mag.*, vol. 43, p. 226 (1897).

length as the original line α but is plane polarized in a plane at right angles to the magnetic field H , the direction of polarization being indicated by the double arrow above the line. The other two components l_R (left, shorter wave length) and r_R (right, longer wave length) are plane polarized in a direction parallel to the magnetic field.

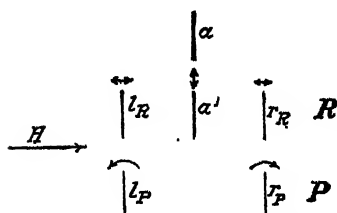


FIG. 15.—Diagrammatic representation of the Zeeman effect: α is the original line; R shows the three components when viewed at right angles to the magnetic field; P shows the two components when viewed parallel to the field.

If the pole pieces of the electromagnet be drilled through longitudinally so that one may view the source in a direction parallel to the magnetic field, only two components l_P and r_P are seen, as shown at P . These two lines have the same respective wave lengths as the outside components in the previous case, but they are *circularly* polarized in

opposite directions, as shown by the arrows above the lines. (In this latter case the magnetic field is directed toward the reader.)

In his original paper announcing his discovery, Zeeman discusses the phenomenon at first in terms of a mechanical ether, but finally he decides in favor of an explanation in terms of the electrical theory of matter, which had recently been developed in more complete form by Lorentz. It may be worth while to derive here the classical Zeeman-Lorentz theory of the effect, in spite of the fact that we now believe classical theory not to be applicable to atomic phenomena; a simple concrete picture is thus obtained which assists the memory, in contrast with the abstractness of the wave-mechanical theory. Furthermore, a motion such as Zeeman supposed to occur in the atom could actually be executed by an electron in a vacuum tube.

In order to have light of a fixed frequency emitted by electrical charges, according to classical theory, the charges must vibrate in simple harmonic motion. Let there be then, within an atom, a particle having a charge of e electromagnetic units and a mass m (e being either positive or negative). When this particle is displaced a distance r from its normal position of equilibrium O , let it be acted upon by a restoring force br directed toward O . Then the component of the force parallel to any fixed line (e.g., AB in Fig. 16) will also equal b times the component of the displacement in this direction. Hence, this component of the displacement will vary harmoni-

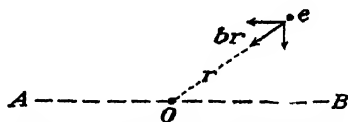


FIG. 16.—A particle vibrating harmonically in two dimensions.

cally with a period

$$T = 2\pi \sqrt{\frac{m}{b}}, \quad (43)$$

according to the ordinary laws of simple harmonic motion. Any component perpendicular to AB will do the same but, perhaps, with different amplitude and phase. Thus, the general motion of the particle will be equivalent to three independent vibrations in mutually perpendicular directions, all executed with the same period T .

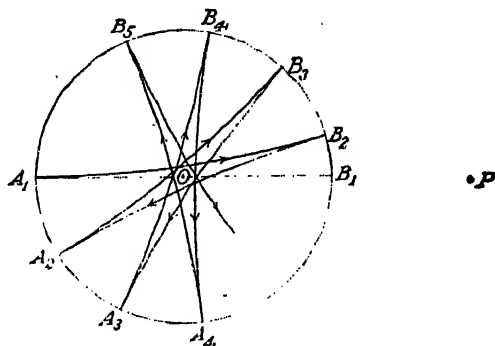


FIG. 17.—Effect of a magnetic field on a charged body vibrating initially with simple harmonic motion at right angles to the field.

Now suppose the particle is vibrating in a single direction, say along the line AB in Fig. 16 or A_1B_1 in Fig. 17. Let a magnetic field of strength H be applied at right angles to the plane of the paper and directed away from the reader. According to the familiar elementary law, the charge will now experience an additional force f_n , at right angles both to the field and to its velocity v , whose magnitude is

$$f_n = Hcv,$$

c being the magnitude of the charge in electromagnetic units. The effect of this force upon the charge will be to deflect it sideways. If it is a positive charge and starts from A_1 toward B_1 (Fig. 17), it will arrive at some such point as B_2 ; starting back toward O , it will be deflected again and arrive at A_2 , and so on. Thus, the field causes the path to rotate slowly, in a plane perpendicular to the field.

The rate of this rotation can be found by resolving the motion into two circular motions about O . Suppose first that there is no magnetic field. Then the particle is capable of revolving around the circle shown in Fig. 17 with a steady speed v_0 or an angular velocity ω_0 and

period of revolution T_0 given by

$$\frac{mv_0^2}{r} = m\omega_0^2 r = br, \quad \omega_0 = \sqrt{\frac{b}{m}} = \frac{2\pi}{T_0}. \quad (44a, b)$$

Suppose now we imagine a motion in which the particle revolves in this manner, and another in which the motion is similar but is counterclockwise, and add the two resulting displacements due to these two motions on the assumption that the particle starts from B_1 at the same time in both. Then components of displacement perpendicular to A_1B_1 cancel out, and we have a linear harmonic vibration of period T along A_1B_1 . What we have done is simply to amplify the treatment of simple harmonic motion as given in elementary texts.

Now introduce a magnetic field, directed, as before, away from the reader in Fig. 17. In addition to the elastic force br , there will then be the force Hev ; if the particle moves at constant speed around the circle in a counterclockwise direction and is positively charged, this added force will be constant and directed toward the center. A steady circular motion is, therefore, still possible, but the speed must be given not by (44a) but by

$$\frac{mv^2}{r} = br + Hev.$$

Putting $v = \omega_1 r$, where ω_1 is the new angular velocity, we have, after rearranging,

$$\omega_1^2 - H \frac{e}{m} \omega_1 = \omega_0^2,$$

for, by Eq. (44b), $b/m = \omega_0^2$. Completing the square and extracting the square root gives

$$\omega_1 - \frac{He}{2m} = \sqrt{\omega_0^2 + \frac{H^2 e^2}{4m^2}}. \quad (45a)$$

If the motion is clockwise, instead of counterclockwise, the force due to the magnetic field acts in the opposite direction and we find

$$\omega_1 + \frac{He}{2m} = \sqrt{\omega_0^2 + \frac{H^2 e^2}{4m^2}}. \quad (45b)$$

In all cases of interest, however, $He/2m$ is very small in comparison with ω_0 , so that the radicals in these equations may be replaced by ω_0 . Hence, if we change to frequencies by writing $\omega_0 = 2\pi\nu_0$, $\omega_1 = 2\pi\nu_1$, and $\omega_2 = 2\pi\nu_2$, we have

$$\nu_1 = \nu_0 + \frac{He}{4\pi m} \text{ (counterclockwise),} \quad (46a)$$

$$\nu_2 = \nu_0 - \frac{He}{4\pi m} \text{ (clockwise).} \quad (46b)$$

If two such circular motions with slightly different frequencies are superposed, it is easily seen that their resultant is a rotating linear vibration such as is illustrated in Fig. 17, and that the quantity $He/4\pi m$ thus gives the rate at which the direction of vibration is caused to precess about the direction of the magnetic field. This effect is known as the "Larmor precession." However the particle may be vibrating, its component of motion *perpendicular* to the field can be resolved into two circular motions of the type described. Its component of motion *parallel* to the field, on the other hand, is unaffected and occurs at the undisturbed frequency ν_0 .

The characteristic features of the emitted radiation are easily inferred now if we recall that the radiated electric intensity is as nearly oppositely directed to the acceleration of the charge as it can be, subject to the condition that it must be perpendicular to the direction of propagation of the waves (see Sec. 37). If the emitted light is viewed with a spectroscope in a direction *perpendicular* to the magnetic field, as by an observer at P in Fig. 17, three lines will be seen. One, emitted by the component of vibration parallel to the field, will have a frequency ν_0 and will be polarized with the electric vector parallel to the applied magnetic field; the two others, emitted by the circular motions, will have frequencies ν_1 and ν_2 and will be polarized with the electric vector perpendicular to the magnetic field. If, on the other hand, the light is viewed in a direction *parallel* to the field, only the lines of frequency ν_1 and ν_2 can be seen [for in Eq. (33) $\sin \theta = 0$ for the other component of the vibration]; and they will be circularly polarized. These features agree exactly with Zeeman's observations.

For a *positive* particle, however, the slower rotation occurs in a clockwise direction as seen by an observer looking in the direction of the field [cf. Eq. (46b)], whereas Zeeman found counterclockwise rotation for the line of lower frequency (cf. Fig. 15, P , where the magnetic field is assumed to be directed *toward* the reader). He concluded, therefore, that the charge on the radiating particles must be *negative*. The ratio e/m for the particles can be computed from the separation between the outer lines. Letting e stand hereafter for the numerical charge on a particle, we have from Eqs. (46a,b)

$$\Delta\nu = \nu_2 - \nu_1 = \frac{He}{2\pi m}, \quad \frac{e}{m} = \frac{2\pi}{H} \Delta\nu. \quad (47a,b)$$

Zeeman's first observation indicated that e/m must be "of the order of magnitude of 10^7 electromagnetic units." In a later experiment, working with much higher resolving power, he found $e/m = 1.6 \times 10^7$, as against the modern value of 1.76×10^7 .

It should be remarked, however, that later experiments have shown the Zeeman effect to be much more complicated than Zeeman at first assumed from his observations with apparatus of relatively low resolving power. In general, many more components are observed than are shown in Fig. 15. The simple theory outlined above fails completely to explain these more complicated "patterns." The quantum theory of the Zeeman effect is discussed in Chap. VIII.

44. The Discovery of the Electron.—Previous to 1897, many studies had been made of that beautiful phenomenon, the discharge of electricity through rarefied gases. Let the discharge from an induction coil or an electrostatic machine pass between the negative terminal *C* (cathode) and the positive terminal *A* (anode) sealed into a glass tube (Fig. 18), which is being exhausted through the side tube *T*. At a very low pressure, there appears around the cathode a dark space, known as the "Crookes dark space," which, with further decrease in

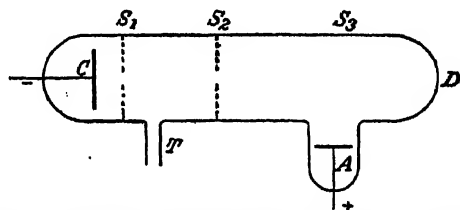


FIG. 18.—Tube for electrical discharge at low pressure.

pressure, grows longer (*i.e.*, extends farther toward *D*), until finally it reaches the glass walls of the tube. The glass is then observed to glow, the color, greenish or bluish, depending on the kind of glass of which the tube is made. If screens pierced with holes *S*₁ and *S*₂ are introduced, the glow is confined to a spot on the end of the tube, at *D*. The "something" which, under these conditions, seems to proceed from the cathode and to cause the phosphorescence of the glass was early called "cathode rays." In support of this view could be cited the usual facts listed in elementary texts of physics. The rays are deflected by a magnetic field; they are also deflected by an electrostatic field, the spot at *D* moving upward if a positively charged rod is brought up in the position *S*₃. The concentration of the spot, in line with the holes in *S*₁ and *S*₂, indicates that something is proceeding from *C* to *D*; if the cathode is suitably curved, this something can be "focused" upon a piece of platinum foil within the tube, which is heated to incandescence if the rays are sufficiently intense. Finally, in 1895, Perrin caught the rays in an insulated chamber connected to an electroscope and proved that they carry a negative charge.

Final confirmation of the correctness of the view that cathode rays are moving negatively charged particles came from the classical experiments¹ of J. J. Thomson. His experiments are so fundamental in the history of the electrical theory of matter as to warrant description. The highly evacuated glass tube [Fig. 19(a)] contains the cathode C and an anode A , which has a small rectangular slot in it through which the cathode rays may pass. B is a screen similar to, and electrically connected with, A . Cathode rays, accelerated from C toward A , after passing the slot in A move with uniform velocity and emerge from the slot in B as a small bundle of rectangular cross section, which causes a small, fluorescent patch at P_1 on the far side of the bulb.

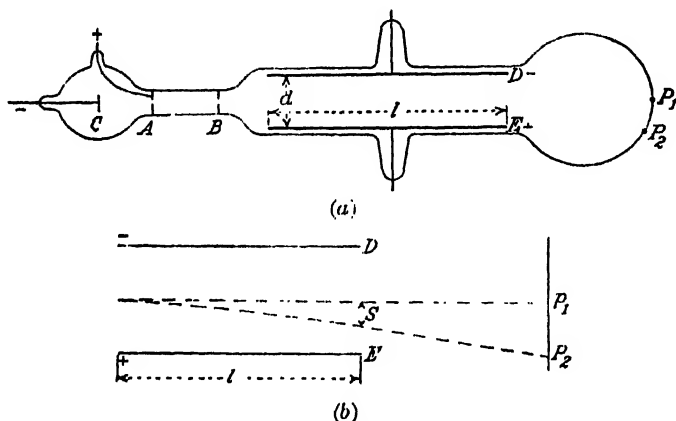


FIG. 19.—(a) Thomson's apparatus for determining the ratio e/m for electrons. (b) Deflection of an electron by the electrostatic field.

But when a potential difference V is maintained between the parallel plates D and E , E being positive, the spot appears at P_2 , having been deflected downward by the electrostatic field. A pair of Helmholtz coils, not shown, whose diameters are equal to the length of the plates D and E , are placed, one in front and the other behind the tube, so as to produce a magnetic field perpendicular to the plane of the paper, the strength of which can be determined from the dimensions of the coils and the current through them. If the magnetic field is directed *toward* the reader, the spot P_1 is deflected *upward*.

Two experiments are now performed:

1. With a given electrostatic field between the plates, the strength of the magnetic field is adjusted to such a value as will cause the spot to return to the original undeviated position P_1 .

¹ *Phil. Mag.*, vol. 44, p. 293 (October, 1897).

2. The magnetic field is then removed and the deflection P_1P_2 caused by the electrostatic field alone is measured.

From these two experiments, the ratio e/m may be determined as follows. In the first experiment, the electrostatic field has a magnitude $E = V/d$, if d denotes the distance between the plates, and it exerts a force Ee upon the charge e . Let both E and e be measured in electromagnetic units. This force is just equal and opposite to the force Hev due to the magnetic field, H being the field in oersteds and v the velocity with which the particles emerge from B . The velocity is constant over the entire path from B to P_1 because there are no longitudinal forces acting on the particle. Hence $Hev = Ee$ and

$$v = \frac{E}{H}. \quad (48)$$

Thus the first experiment serves to measure the velocity of the particles as they emerge from B . In the second experiment, let the distance P_1P_2 be measured; from this, the deflection S of the particles as they pass over the distance l between the plates D and E can be determined [cf. Fig. 19(b)]. This deflection results from a uniform acceleration

$$a = \frac{Ee}{m},$$

acting during a time l/v ; hence, by the laws of uniformly accelerated motion,

$$S = \frac{1}{2} E \frac{e}{m} \left(\frac{l}{v} \right)^2. \quad (49)$$

All quantities in this equation being known except e/m , the latter may be computed.

Thomson found that the value of e/m determined in this way was of the order of 10^7 and that *it was independent of the kind of gas in the tube* (air, H_2 , or CO_2) and, likewise, independent of the material of the electrodes (Al, Fe, or Pt). A later determination gave

$$\frac{e}{m} = 1.7 \times 10^7,$$

a value almost numerically identical with the value of e/m determined from the Zeeman effect for the particles taking part in light emission. (The modern value is 1.76×10^7 .) This value of e/m is very much larger than the value of e/m for hydrogen atoms in electrolysis. Its great magnitude might result either from a large value of e or a small

value of m or both. It became a matter of much importance, therefore, to determine the charge e carried by these particles.

Now it seemed almost certain that the charge on the cathode particles must be of the same order as the charges on gaseous ions such as are produced in a gas by X-rays. The study of such ions had begun just a few years before Thomson's work. To measure their charges, Townsend, working in Thomson's laboratory, utilized the clouds that form about the ions in saturated air. By observing the rate of fall of the cloud and applying Stokes' law for the free fall of spheres through a viscous medium, he was able to determine the size of the droplets; from a measurement of the total amount of water in the cloud, he could then calculate the number of droplets it had contained. He assumed that each droplet contained just 1 ion; hence, having measured also the total charge on the cloud, he was able to calculate the charge on a single ion. For this charge he obtained a value of about 3×10^{-10} electrostatic unit. A repetition of the measurements with some modifications by Thomson gave the value 6.5×10^{-10} electrostatic unit.¹

Thomson assumed that the charge on the gaseous ions was the same as that on his cathode particles. It followed, then, that the cathode particles must be previously unknown particles of extremely small mass, which he called "corpuscles" or "primordial atoms." For many years English writers stuck to the name "corpuscle" for these particles, using the word "electron" in Stoney's original sense to denote the amount of charge carried by a corpuscle or a univalent ion; but others, including Lorentz, called the corpuscles themselves "electrons," and this usage ultimately became well established. It was generally assumed that electrons are a constituent part of all atoms and are responsible for the emission of light by them, thus accounting for the fact that the ratio e/m had been found to be the same for the vibrating particles causing the Zeeman effect as for the cathode rays. Thus was made possible the explanation of a number of more or less diverse phenomena on the basis of a single concept.

45. Electronic Magnitudes.—A much more reliable method of measuring ionic charges was developed by Millikan in 1909. He found that tiny droplets of oil in ionized air, viewed with a microscope, would frequently pick up charges and could then be held suspended, or accelerated upward or downward, by applying a suitable electric field. When uncharged, the droplets fall at a slow uniform rate, their weight being balanced by the drag due to the viscosity of the surrounding air.

¹ For a good account of work on the electronic charge see Millikan, "Electrons (+ and -), Protons, Photons, Neutrons, and Cosmic Rays," 1935.

By observing their rate of fall, Millikan was able to determine the size and weight of the droplets. According to Stokes' law, a sphere of radius a moving at a steady slow speed v through a fluid whose coefficient of viscosity is η experiences a resisting force

$$F = 6\pi\eta av_0. \quad (50)$$

If we equate this force to the weight of a droplet, which is $\frac{4}{3}\pi a^3 \rho g$ in terms of the density ρ of the oil, we have

$$\frac{4}{3}\pi a^3 \rho g = 6\pi\eta av_0. \quad (51)$$

Now suppose the drop picks up a charge of e units and that a vertical electric field E is present, e and E both being measured in electrostatic units. Then a force eE is added to the weight and instead of (51) we have

$$\frac{4}{3}\pi a^3 \rho g + eE = 6\pi\eta av_1,$$

v_1 being the new velocity of steady fall. From these equations we find

$$eE = 6\pi\eta a(v_1 - v_0);$$

and, inserting in this equation the value of a given by (51) and solving for e , we find

$$e = 6\pi\eta^{3/2}(v_1 - v_0) \left(\frac{9v_0}{2\rho g}\right)^{1/2} \frac{1}{E} \quad (52)$$

All quantities in this last expression being known, e can be calculated.

Millikan found that the charges so calculated from his observations were all multiples of a smallest charge. The latter he assumed to be the charge on the electron, an assumption which is difficult to test directly but seems to be well supported by indirect evidence. For the electronic charge he found $e = 4.774 \times 10^{-10}$ electrostatic unit.

This latter value of e was accepted for many years. Its accuracy was first called into question by Bäcklin,¹ who, from values of X-ray wave lengths measured by a ruled grating found $e = 4.793 \times 10^{-10}$ e.s.u., a value 0.4 per cent higher than that reported by Millikan. It now appears that the discrepancy was due to Millikan's having used too low a value of the viscosity of air, which appears as η in Eq. (52) above. Recent measurements give a value of η about 0.5 per cent higher than Millikan's,² and this higher value leads to an "oil-drop value" of e of $(4.8036 \pm 0.0048) \times 10^{-10}$ e.s.u. Dunnington concludes that the best

¹ BÄCKLIN, E., "Absolute Wellenbestimmungen der Röntgenstrahlen," Dissertation, Upsala, 1928.

² Cf. DUNNINGTON, *Rev. Modern Physics*, vol. 11, p. 71 (1939).

"ruled-grating" value of e is $(4.8025 \pm 0.0004) \times 10^{-10}$ e.s.u. The best modern value of the ratio e/m , which has been measured recently by different methods with high precision, appears to be¹ 1.7591×10^7 . From e and e/m , the value of the electronic mass m can then be calculated.²

We thus obtain for these three important constants and for N_0 , representing Avogadro's number or the number of atoms in a gram-atom, calculated as F/c where F is the value of the faraday (Sec. 42):

$$\begin{aligned} e &= 4.803 \times 10^{-10} \text{ electrostatic unit,} \\ \frac{e}{m} &= 1.7591 \times 10^7 \text{ electromagnetic units,} \\ m &= 0.9107 \times 10^{-27} \text{ gram.} \\ N_0 &= 6.023 \times 10^{23}. \end{aligned}$$

It is often of interest, also, to know how the electron compares in mass with a hydrogen atom. The faraday, mentioned in Sec. 42, represents the charge carried by 1 gram-atom or 1.0078 grams of hydrogen; its value is 9,648.9 electromagnetic units. Thus for a hydrogen atom of mass M_H , which carries the same charge as the electron when ionized, $e/M_H = 9,648.9/1.0078$ or

$$\frac{e}{M_H} = 9,574 \times 10^3 \text{ electromagnetic units.}$$

Dividing this number into the value of e/m for the electron, we have, e canceling,

$$\frac{M_H}{m} = 1,837$$

as the ratio of the two masses.

46. Photoelectrons.—We now return to the question raised at the end of Sec. 41: What are the carriers of electricity in the photoelectric current? The discovery of the electron furnished an answer: *The photoelectric effect is due to the liberation, from the illuminated metal plate, of electrons which, under the influence of the electric field, pass from cathode to anode, thereby causing the photoelectric current.* This hypothesis

¹ Cf. BIRGE, *Phys. Rev.*, vol. 51, p. 972 (1938); DUNNINGTON, *loc. cit.*

² This value of m is the so-called "rest mass" of the electron, *i.e.*, its mass when at rest or when moving with velocities negligibly small compared with the velocity of light. If we denote this value by m_0 , the mass m , of the electron when moving with velocity v is (cf. Sec. 67)

$$m_v = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

was confirmed by Lenard,¹ who showed that the photoelectric discharge is deflected in a magnetic field exactly as are cathode rays. By measuring the deflection of the "photoelectric rays" in a known magnetic field, he found a value of e/m of (about) 1.2×10^7 , in qualitative agreement with Thomson's value of e/m for electrons.²

Lenard's method of determining e/m for photoelectrons involves basic principles which, with ever increasing refinement, have been widely applied in "charged-particle" physics. His apparatus is shown diagrammatically in Fig. 20. A glass tube, which could be exhausted to the highest attainable vacuum through the side tube T , contained an aluminum cathode C , which could be illuminated by ultraviolet light from a spark S , the light passing through the quartz plate Q . The

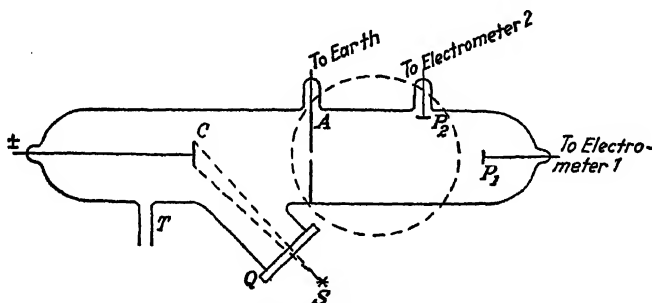


FIG. 20.—Lenard's apparatus for determining the ratio e/m for photoelectrons.

cathode C could be charged to any potential, positive or negative. A screen A , with a small hole at its center and connected to earth, served as anode. P_1 and P_2 were small metal electrodes connected to electrometers. When C was illuminated and charged to a negative potential of several volts, photoelectrons were liberated and accelerated toward the anode A . A few electrons passed through the hole in the center of A and proceeded thereafter at uniform velocity to the electrode P_1 , their reception there being indicated by the electrometer 1. But if, by means of a pair of Helmholtz coils (represented by the dotted circle), a magnetic field directed toward the reader was produced in the region between A and P_1 , the electrons were deflected upward in a circular path and, with a sufficient field strength, struck the electrode P_2 .

Lenard first investigated the relation between the current reaching the anode and the potential V applied to C , A being assumed to be

¹ *Ann. d. Physik*, vol. 2, p. 359 (1900).

² Later, Alberti [*Ann. d. Physik*, vol. 39, p. 1133 (1912)] found for photoelectrons $e/m = 1.765 \times 10^7$ e.m.u. gram⁻¹.

always at zero potential. There was no photoelectric current when V was several volts positive. But, when V was dropped to about 2 volts *positive*, a small current was observed. This indicated that the photoelectrons were not simply *freed* from the cathode but that some of them at least *were ejected with sufficient velocity to enable them to overcome the retarding potential of 2 volts*. The current increased when V was reduced to zero and increased still more rapidly as V was made negative but attained a "saturation" value after V had reached some 15 or 20 volts negative. These data are shown diagrammatically in Fig. 21. I is the photoelectric current and V_0 is the positive potential (about 2 volts) necessary to apply to the cathode to prevent the escape of any electrons.

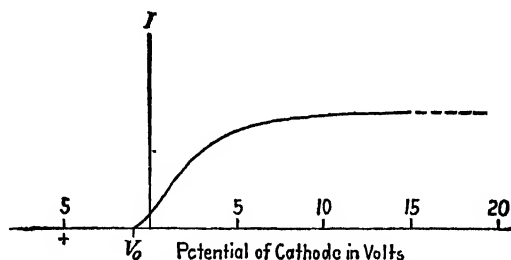


FIG. 21.—Variation of photoelectric current with cathode potential.

The determination of e/m was made essentially as follows. Let a negative potential V , large compared with V_0 , be applied to the cathode C , all potentials being in electromagnetic units. The photoelectron, on reaching the anode, will have a kinetic energy given approximately by

$$Ve = \frac{1}{2}mv^2,$$

where m and e are the mass and the charge in electromagnetic units of the electron, respectively, and v is its velocity on reaching A (Fig. 20). Assuming that after leaving A the electron moves in a uniform magnetic field, the circular path that it follows is determined by the equation

$$Hev = \frac{mv^2}{R},$$

where H is the strength of the magnetic field just necessary to cause the electron to reach P_2 , and R is the radius of the corresponding circular path, determined from the geometry of the apparatus. From the last two equations we have

$$\frac{e}{m} = \frac{2V}{H^2 R^2},$$

from which e/m may be computed.

47. Relation between Photoelectric Current and Intensity of Illumination of the Cathode.—The experiments of Elster and Geitel, Lenard, and Ladenburg seemed to show that, as long as there is no change in the spectral quality of the light causing the emission of electrons, the photoelectric current is apparently proportional to the intensity of illumination on the emitting surface, although some work by Griffith threw doubt on this conclusion. Subsequent experiments,¹ however, confirmed the law of proportionality and showed that it holds rigorously over a very wide range² of intensities—in the experiments of Elster and Geitel, over a range of 50,000,000 to 1. This law of proportionality was, therefore, well established, and it is one of the most important laws of the photoelectric effect.

48. Energy Distribution of Photoelectrons.—In his experiments, Lenard showed, by observing the retarding potential against which these electrons could move, that for a given emitter their kinetic energy did not exceed a definite maximum, given by V_0e , where V_0 is the maximum retarding potential for which a photoelectric current was observed and e is the electronic charge. From Lenard's time to the present, this question of the energy of photoelectrons has been the subject of many investigations.

We mention the experiments of Richardson and Compton,³ who, by introducing important corrections and techniques, cleared up a number

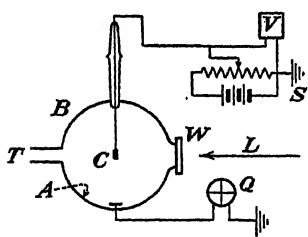


FIG. 22.—Apparatus of Richardson and Compton for observations on the energy of photoelectrons.

of previous discrepancies and gave impetus to the quantitative verification of Einstein's photoelectric equation (see next section). The apparatus used by Richardson and Compton is sketched in Fig. 22. The photoelectric emitter C , a strip 1×5 mm. of the metal under investigation, was placed at the center of a spherical glass bulb B some 7.5 cm. in diameter and silvered on its inner surface. The emitter C could be charged to any desired potential, read by voltmeter V . The bulb could be evacuated through the side tube T . Monochromatic light L passed through the quartz window W and fell on C . The silver coating on the inside of B , serving as anode A , was connected to the electrometer Q by which the photoelectric current was measured.

¹ RICHTMYER, *Phys. Rev.*, vol. 29, pp. 71, 404 (1909).

² ELSTER and GEITEL, *Phys. Zeits.*, vol. 14, p. 741 (1913); vol. 15, p. 610 (1914); vol. 17, p. 268 (1916). KUNZ and STEBBINS, *Phys. Rev.*, vol. 7, p. 62 (1916).

³ RICHARDSON and COMPTON, *Phil. Mag.*, vol. 24, p. 575 (1912).

The large spherical anode, with the small emitter at its center, served two important purposes. First, since the electric field around C was nearly radial, it was possible to measure the energy distribution of the photoelectrons irrespective of direction of emission and thus to determine what has come to be known as the "total energy," in contradistinction to the "normal energy," which is measured when cathode and anode are a pair of parallel plates, as in Fig. 14. Second, the impact of the photoelectrons on the anode caused the diffuse emission from the latter of a certain number of secondary electrons. In the case of parallel plates, a considerable proportion of these return to the cathode, and the observed current is, therefore, not the true photoelectric current. Only a few of these secondaries, however, reach the small cathode C at the center of the large anode, and the error in the observed current is negligible. Richardson and Compton made careful correction for the contact difference of potential between the cathode C and the silver anode, an important correction, for the maximum retarding potentials 1 or 2 volts—were of the same order of magnitude as the contact potentials.

The relation between photoelectric current and retarding potential obtained by Richardson and Compton is shown for aluminum as emitter in Fig. 23. When the aluminum surface was illuminated by ultraviolet light of constant intensity and of wave length 0.000020 cm., no photoelectric current was observed when the potential of C was greater than 2.3 volts— V_5 , curve 5. As the retarding potential decreased from this value to zero, the current rose to a

maximum I_m , beyond which there was no increase for negative potentials on the cathode. For longer wave lengths, similar curves—1, 3, 2, 1—were obtained, but with the important difference¹ that the critical retarding potentials— V_4 , V_3 , V_2 , V_1 —became progressively less as the wave length increased. Curves of the same shape were obtained for other

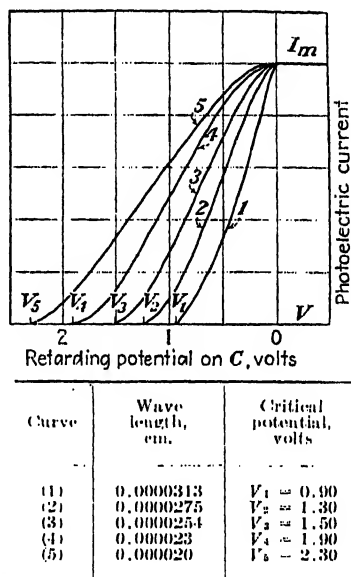


FIG. 23. Variation of photoelectric current with retarding potential.

¹ The intensity of these several monochromatic beams was adjusted to give the same value of I_m in each case.

intensities of illumination, the current increasing in all cases in proportion to the intensity of the light.

These observations are readily interpreted on the assumption that, for a given emitter, light of wave length λ causes the emission of photoelectrons with initial velocities v varying from zero up to a maximum v_m determined by λ . The typical variation of photoelectric current with retarding potential V is shown again in Fig. 24. Since no

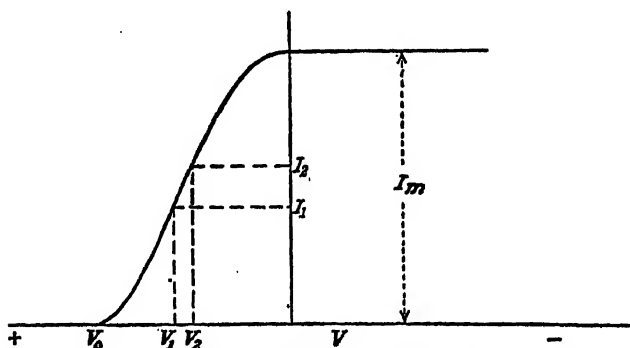


FIG. 24.—Typical variation of photoelectric current with retarding potential.

photoelectric current was observed for retarding potentials greater than V_0 , v_m is at once given by

$$V_0 e = \frac{1}{2} m v_m^2,$$

e being the numerical electronic charge. When the potential V applied to the emitter is zero or negative, all electrons freed by the light reach the anode and the photoelectric current has its maximum value I_m . When $V_0 > V > 0$, only those electrons will reach the anode for which

$$\frac{1}{2} m v^2 > V e.$$

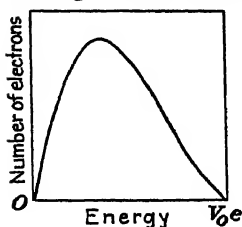


FIG. 25.—Distribution of photoelectrons as to energy.

Let the photoelectric currents be I_1 and I_2 when the potentials are V_1 and V_2 , respectively. Then $(I_2 - I_1)/e$ photoelectrons per second leave the emitter with velocities between v_1 and v_2 defined by

$$V_1 e = \frac{1}{2} m v_1^2 = E_1, \quad V_2 e = \frac{1}{2} m v_2^2 = E_2,$$

E_1 and E_2 being the corresponding kinetic energies. Thus the slope of the curve in Fig. 24 at any point is proportional to the number of photoelectrons possessing energy E corresponding to the value of V at that point. A curve plotted between these slopes and V is the

energy distribution curve for photoelectrons; its general form is shown in Fig. 25.

49. Relation between the Velocities of Photoelectrons and the Frequency of the Light.—As stated above, if the photoelectric current as a function of applied voltage V is determined for several different monochromatic radiations, curves of the type shown in Fig. 23 are obtained. $V_1, V_2, \dots V_5$ denote the various values of the critical retarding voltage necessary to prevent the escape of the fastest electron. We note that, as the wave length decreases, the critical retarding voltage increases. Or, *the maximum kinetic energy of the photoelectrons increases with increasing frequency ν of the light that causes their emission.*

A very simple linear relation has been found to exist between this maximum energy and the frequency. If, as shown by Millikan, a curve is plotted between $V_0 c$ and the corresponding frequency ν of the light, a straight line results which has an intercept ν_0 on the frequency axis (Fig. 26). The experimental meaning of this intercept is that light of frequency less than ν_0 cannot cause the emission of photoelectrons from the metal concerned. The quantity ν_0 is characteristic of the emitting electrode, but *the slope of the curve is the same for all electrodes.* The equation of the curve may be written

$$V_0 c = \frac{1}{2} m v_m^2 = h(\nu - \nu_0) = h\nu - h\nu_0$$

or

$$\frac{1}{2} m v_m^2 = h\nu - \omega_0, \quad (53)$$

where h is the slope of the curve and ω_0 is written for $h\nu_0$. If, in Fig. 26 and Eq. (53), the ordinates are expressed in ergs and the abscissas in frequencies (sec.^{-1}), the numerical value of h , as found by Millikan, is 6.56×10^{-27} erg sec. The best modern value is 6.61×10^{-27} . This constant h , called "Planck's constant" for reasons which will appear in Chap. V, is one of the fundamental constants of nature and has played an extraordinary part in modern physics. *The product $h\nu$ is called a quantum of energy corresponding to light of frequency ν .*

Equation (53) has had a very interesting history and is one of the most fundamental equations, or laws, of modern physics. It was first proposed on theoretical grounds by Einstein, in 1905, as a result of the extension to the photoelectric process of the concept, previously developed by Planck (see Chap. V), that interchanges of energy

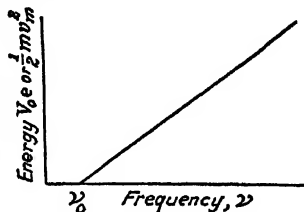


FIG. 26.—Variation of maximum energy of photoelectrons with frequency of the exciting radiation.

between radiation and matter take place in energy quanta $h\nu$, where ν is the frequency of the radiation absorbed or emitted and h is a constant. Einstein arrived at the equation¹ by assuming that the whole quantum $h\nu$ of radiant energy was absorbed by an electron but that a part ω_0 of it was expended by the electron in escaping from the emitter. He had at his disposal only qualitative data to show that his equation gave results of at least the right order of magnitude. The equation received final and complete experimental verification as a result of the precision experiments of Millikan, to which reference has been made previously.² Subsequent work extended the validity of the equation to the X-ray region, where frequencies are involved that are several thousand times the frequencies of visible light. Einstein's photoelectric equation played an enormous part in the development of the modern quantum theory. But in spite of its generality and of the many successful applications that have been made of it in recent theories in physics, the equation is, as we shall see presently, based on a concept of radiation—the concept of “light quanta”—completely at variance with the most fundamental postulates and conclusions of the electromagnetic theory of radiation.

50. Other Properties of Photoelectric Emission.—It is not our purpose to give a complete account of the properties of the photoelectric effect, which has been extensively studied during the past two decades. A few other features of interest may be mentioned, however, in passing.

We have seen that monochromatic light causes the emission of photoelectrons only if its frequency lies above a certain critical value ν_0 , commonly called the “threshold” frequency. This critical frequency varies considerably with the state of the surface, as does the photoelectric current in general. Usually it lies in the ultraviolet, but for the alkali metals and for barium and strontium it lies in the visible region. For potassium the threshold lies in the red, and for cesium even in the infrared. Photoelectric cells used for the measurement of light commonly contain either an alkali metal or barium.

Variation of temperature usually has little or no effect upon the photoelectric current, so long as the temperature does not exceed several hundred degrees centigrade and so long as no change occurs in the crystal structure or in the physical state of the metal. The alkalis, however, form an exception to this rule.

If *polarized light* is used, differences are commonly found as the plane of polarization is rotated, except, of course, at normal incidence.

¹ EINSTEIN, *Ann. d. Physik*, vol. 17, p. 132 (1905).

² MILLIKAN, *Phys. Rev.*, vol. 7, p. 355 (1916).

The effect is complicated, and its cause is in doubt. Especially interesting is the "selective effect" in the alkali metals. Over certain ranges of wave length, the photoelectric current from these metals is much greater when the electric vector in the light has a component perpendicular to the surface than when it is parallel to the surface; in the case of certain sodium-potassium alloys, the ratio of the currents in the two cases may be 10:1 or 20:1 or even more.

For further discussion of the photoelectric effect the student is referred to the treatise of Hughes and DuBridge.¹

We shall return here to those fundamental problems presented by the photoelectric effect which have a bearing on quantum theory.

51. Source of the Photoelectric Energy.—Whence comes the energy which the photoelectrons are observed to carry away from the surface of the emitter? Does it come from energy previously stored in the atom, or by direct absorption from the incident light? And, in the latter case, what is the mechanism by which the energy is absorbed? As a basis for the discussion of this question, we have the following experimental conclusions, previously discussed:

1. For a given spectral composition of the incident light, the number of photoelectrons expelled per second from a given emitter is strictly proportional to the intensity of the light over a very wide range of intensities.

2. The maximum kinetic energy of the photoelectrons is a linear function of the frequency of the light, the relation between the two being given by the Einstein photoelectric equation

$$\frac{1}{2}mv_m^2 = h\nu - \omega_0.$$

If we interpret ω_0 as the amount of work required to escape from the emitter, it follows that each electron receives from the light an amount of energy $h\nu$ which is strictly proportional to the frequency of the light *and is independent of the nature of the emitter.*

3. The maximum kinetic energy of the photoelectrons emitted by monochromatic light has been found to be *rigorously independent* of the intensity of the light over a very wide range of intensities.

From these experimental conclusions, the presumption is very strong that the energy acquired by the photoelectrons is obtained from the incident light. If that is the case, then it must be that the atom absorbs the energy from the light and gives this energy to the photoelectron. But how can we picture an atomic mechanism that will function in accordance with the experimental facts? The difficulties in the way of postulating such a mechanism on the basis of the

¹ HUGHES, A. L., and L. A. DUBRIDGE, "Photoelectric Phenomena," 1932.

classical electromagnetic wave theory of light can best be shown by some numerical data.

A photoelectric current of 1×10^{-12} e.m.u. per square centimeter is easily obtained from a sodium surface *in vacuo* with very moderate intensity of illumination. This current corresponds to the emission of about 6.25×10^7 electrons per square centimeter per second. These electrons come from the surface layers of atoms. We can easily compute the order of magnitude of the number of atoms involved. Calling Avogadro's number N_0 , the atomic weight of the emitter A , and its density δ , the number of atoms per unit volume n is given by

$$n = \frac{N_0}{A} \delta.$$

$N_0 = 6.02 \times 10^{23}$; and, for sodium, $A = 23.0$ and $\delta = 0.97$ gram per cubic centimeter. Therefore, $n = 25.4 \times 10^{21}$ atoms per cubic centimeter. Assuming the atoms in a centimeter cube of sodium to be arranged in regular rows, columns, and layers,¹ we find that there are $\sqrt[3]{25.4 \times 10^{21}} = 2.94 \times 10^7$ atoms along each edge, or $(2.94 \times 10^7)^2 = 8.64 \times 10^{14}$ atoms in a layer 1 cm.² in area and 1 atom deep. Assuming, for the sake of illustration, that the photoelectrons are supplied by, say, the first 10 layers of atoms, we find that

$$8.64 \times 10^{15} \text{ atoms}$$

furnish, under the conditions of illumination above mentioned, 6.25×10^7 electrons per second. That is, 1 atom in

$$\frac{8.64 \times 10^{15}}{6.25 \times 10^7} = 1.38 \times 10^8,$$

on the average, furnishes 1 photoelectron per second. But what is it that determines *which one* of these hundred million odd atoms shall furnish an electron in any particular second? According to the wave theory of light, radiant energy should be distributed continuously and uniformly over the wave front. We know of no reason to expect that one atom differs from another in any way which might result in one atom having at any instant a greater absorbing power than another. *On the basis of the classical theory of radiation*, we are unable to put forward any hypothesis that should result in singling out the one atom in a hundred million.

Furthermore, it is hard to understand how the energy carried off by a photoelectron comes to be concentrated in one spot. For example,

¹ The crystal structure of sodium is really a body-centered cube. See "International Critical Tables," vol. I, p. 340.

an illumination of 0.1 meter-candle, or perhaps 10 ergs per cm.² per sec., produces an easily measurable photoelectric current from sodium. Not over 0.05 erg of this, however, falls within the spectral range that is effective for sodium, which is sensitive only in the blue and near-ultraviolet part of the spectrum. Furthermore, sodium has a high reflecting power, and only *absorbed* light can possibly be used in expelling electrons. We conclude that not over 0.001 erg per cm.² per sec. is available for the photoelectric process.

According to the electromagnetic theory, this absorbed energy should be equally distributed among the 8.6×10^{15} atoms which form the (10, arbitrarily assumed) surface layers, so that there is 1.1×10^{-19} erg per sec. for each atom. Now a quantum of violet light corresponding to $\lambda = 0.00004$ cm. is

$$h\nu = 6.61 \times 10^{-27} \times 0.75 \times 10^{15} = 5 \times 10^{-12} \text{ erg.}$$

Thus, if an atom can absorb only the energy that falls on it (so to speak), the number of seconds required for it to obtain enough energy to expel a photoelectron is

$$\frac{5 \times 10^{-12}}{1.1 \times 10^{-19}} = 4.5 \times 10^7 \text{ seconds.}$$

This is over 500 days!

This calculation may not be quite correct, however. Theoretically, it can be shown that an electron vibrating in *resonance* with the light can absorb as much energy as falls on a considerable fraction of a *square wave length*. There are difficulties about assuming photoelectric emission to be a resonance phenomenon, however; and it is hard to see why there should be a resonator actively absorbing in one atom while all of the neighboring atoms are quiescent. If, however, we brush aside these objections, we may conclude that, in the present case, an atom might possibly absorb the energy falling on 10^{-10} cm., or $10^{-10} \times 10^{-3} = 10^{-13}$ erg per sec. The time required to absorb a quantum would then be only $5 \times 10^{-12} / 10^{-13} = 50$ seconds. Even this, however, is too long a time to agree with experiment. There has never been observed any time lag whatever between the beginning of illumination and the starting of the photoelectric current. Indeed, precise measurements have shown that if there is such a time lag, it is less than 3 billionths of a second!

The classical electromagnetic theory of light thus encounters enormous difficulties in explaining the observed facts in regard to photoelectric emission.

52. The Photoelectric Effect and the Corpuscular Theory of Light.

If we could return to the corpuscular theory and regard light as a rain of "corpuscles" or quanta of energy, all of the difficulties raised in the preceding section would disappear. If we could assume (a) that each quantum, or, in modern terms, "photon," possesses energy $h\nu$ and (b) that a collision between a photon and an atom may result in the absorption of the whole photon and the emission of a photoelectron with the initial energy $h\nu$, we could at once explain why only one atom in many millions expels an electron in any particular second and also why there is no time lag in the photoelectric process.

Such a theory would predict, also, that the photoelectric current should be proportional to light intensity, as is experimentally observed. For the intensity of the light should be proportional to the number of photons falling per second on each square centimeter of the photoelectric emitter; and, because of the large number of photons and atoms involved, the proportion of collisions between photons and atoms which result in the emission of an electron should be constant. Further, we should have a ready explanation of the experimentally observed facts: (1) that the initial velocity of the photoelectrons is independent of the intensity of the light and (2) that this velocity depends only on the wave length of the light. For, keeping the wave length of the incident light constant and increasing the intensity would increase the *number* of photons striking each square centimeter each second and, proportionately, the number of collisions resulting in photoelectric emission; but there would be no change in the *nature of each collision*. Each photoelectron would receive *exactly* the same energy, $h\nu$, and would acquire the same initial velocity regardless of the intensity. If the *frequency* of the incident light is increased, however, the value of $h\nu$ is increased, and, accordingly, the energy received by each photoelectron is increased.

The difficulties with such a radical theory of light, however, are many. For one thing, if we regard light as a "shower" of photons, what can possibly be the meaning of *frequency*? There is nothing periodic about a falling raindrop, for example. As a matter of fact, in order to find the frequency of a beam of light, we *measure* (a) the velocity c of the light and (b) its wave length λ *on the assumption that light consists of waves*, and then we *compute* the frequency as

$$\nu = \frac{c}{\lambda}.$$

Thus we have to rely on the wave theory of light to give us the energy value $h\nu$ of a photon! And there still remains the phenomenon of inter-

ference, which, since its discovery by Young in 1802, has defied explanation on any other basis than by assuming light to be a wave phenomenon. However, the experimental facts of photoelectricity are equally as cogent as the phenomena of interference, and these *cannot be explained on the basis of the classical wave theory of light*.

Thus in 1920, say, the physicist could sum up the situation about like this: On one side of a seemingly impenetrable barrier, or fence, is to be found a group of phenomena—such as interference, polarization, smaller velocity of light in optically denser bodies, indeed, the whole electromagnetic theory and its ramifications—according to which we should say, without the slightest doubt, that light *must consist of waves*. On the other side of the fence is to be found another group—the photoelectric effect, and other phenomena which we shall consider in subsequent chapters—according to which we should say, again without the slightest hesitation, if we did not know what was on the first side of the fence, that light *must be corpuscular*. The situation thus created was perhaps the most puzzling one that has ever arisen in the whole history of physics. We cannot resolve the mystery at this point. We shall return to the question after we have become familiar with the quantum theory of atomic structure and of spectra.

The remainder of the present chapter will be devoted to another aspect of the photoelectric problem, and to a different but related phenomenon involving electrons. Granted that the photoelectrons receive their energy in some manner from the incident light, do these electrons themselves come out of the atoms, or do they come from the so-called “free” electrons in the metal? Before attempting to answer this question, we shall discuss the *spontaneous* emission of electrons from hot bodies.

53. Thermionic Emission.—It has been known for 200 years that air in the neighborhood of hot solids has the power of conducting electricity. In a systematic investigation begun about 1880, Elster and Geitel showed that, in general, at a red heat, charged bodies lose a positive charge more readily than they do a negative one, whereas at a white heat a negative charge is more readily conducted away; a few substances, however, lose a negative charge most readily at all temperatures. About the same time, Edison noticed that a current can flow in an evacuated bulb from a glowing carbon filament to another filament when the hot filament is charged negatively but not when it is positive.

The ions that carry the electricity in such cases were studied by McClelland and by J. J. Thomson. In 1899, Thomson measured e/m , the ratio of charge to mass, for the *negative* ions, using the method

described in Sec. 44. He found this ratio to have about the same large value as for cathode rays and concluded that these negative ions were electrons. Later he showed that, for the positive ions, e/m was in all cases of the same order of magnitude as for electrolytic ions and concluded that the positive ions were positively charged atoms or molecules.

This emission of electrons from hot bodies furnishes the basis for the thermionic vacuum tubes that have come into such wide use. No general treatment of thermionic emission will be attempted here¹; we shall discuss only a few aspects of the phenomenon that are important in connection with the theory of the emission of electrons from conductors.

The thermionic current from a given emitter is found to increase very rapidly with increasing temperature. It depends also upon geometrical factors, which determine space-charge effects, and upon the potential of the emitter relative to its surroundings. From a given emitter operating at a given temperature, however, the thermionic current cannot be made to exceed a certain limiting value. Special interest attaches to this maximum possible thermionic current. Let I denote the maximum current obtainable per square centimeter of the surface from a given emitter operating at absolute temperature T . By means of thermodynamic reasoning, it can be shown that I should vary with temperature, approximately, at least, according to Richardson's equation

$$I = AT^2 \epsilon^{-\frac{e\phi}{kT}}. \quad (54)$$

The proof is rather elaborate and will not be given here;² the equation itself can be understood without going through its deduction. The Napierian base of logarithms is denoted here by ϵ to distinguish it from the numerical electronic charge e ; k is the Boltzmann constant, or gas constant for 1 molecule (1.381×10^{-16} erg per degree centigrade), and ϕ and A are constants depending on the emitting substance and on the state of its surface.

The Richardson equation, as stated in (54) above, is found to represent the experimental facts very well. The variation of the exponential factor in the equation is so rapid, however, that it is not possible to tell whether the factor T^2 is really correct. An argument from quantum theory indicates that the constant A ought, perhaps, to be a universal constant, equal to 60 amperes per square centimeter for

¹ Cf. REIMANN, A. L., "Thermionic Emission," 1934.

² See RICHARDSON, O. W., "Electron Theory of Matter," Chap. XVIII, 1916; REIMANN, *loc. cit.*

all substances; but experimentally it is found to vary over a considerable range from one substance to another.

The quantity $e\varphi$ in the Richardson equation represents an energy. For this reason, φ , known as the "work function," is commonly measured in volts, representing the change of potential through which the electron must pass in order to gain or lose $e\varphi$ ergs of energy. In the theoretical derivation of the equation, $e\varphi$ enters as representing the "heat of emission" of the thermions at the absolute zero of temperature, or, in other words, it is the amount of energy required to extract a thermion at that temperature. The value of φ can be calculated from observations of I by plotting $\log (I/T^2)$ against $1/T$; according to Richardson's equation

$$\log \frac{I}{T^2} = \log A - \frac{e\varphi}{k} \frac{1}{T}$$

so that a straight line is obtained whose slope is $-e\varphi/k$, φ being in electrostatic units of potential. Values of φ so obtained, when converted into volts, range from 2 to 6 volts for various metals.

54. Relation between Thermionic and Photoelectric Constants.—It is obvious that the photoelectric effect and the thermionic emission of electrons are in some respects similar phenomena. They seem to differ mainly in the manner in which the electron is given the additional energy to enable it to escape through the surface of the emitter against the potential barrier. In the thermionic case, the electron acquires this energy from the thermal agitation incident to the high temperature of the emitter; in the photoelectric case, by the absorption of light. In the case of thermionic emission, it is natural to suppose that the heat of emission $e\varphi$ represents work done by the electron in escaping from the emitter, just as the larger part of the heat of evaporation of a liquid represents work done by the evaporating molecules against the attraction of the molecules left behind in the liquid. If that is so, $e\varphi$ ought to be the same as the quantity ω_0 or $h\nu_0$ in the Einstein photoelectric equation, which was assumed to represent work done by the photoelectron. Thus, the thermionic work function ought to be related to the photoelectric threshold, the ν_0 of Fig. 26, by the equation

$$\varphi = \omega_0 = \frac{h\nu_0}{e}, \quad (55)$$

φ and e being measured here in a consistent set of absolute units. Since φ and ν_0 can both be determined experimentally, a check of this relationship is possible.

It has not been easy to eliminate the many spurious disturbances inherent in both photoelectric and thermionic research. But reasonably comparable measurements have been made on several metals with results¹ as follows:

Metal	$h\nu_0/e$, volts	ϕ , volts
Platinum.....	6.30	6.27
Rhodium.....	4.57	4.58
Tantalum.....	4.05	4.07
Tungsten.....	4.58	4.52
Palladium.....	4.96	4.99
Silver.....	4.73	4.08
Gold.....	4.82	4.42

The agreement between the values of the work function determined photoelectrically and those determined thermionically seems good enough to warrant the hypothesis that the photoelectrons and the thermionic electrons have a common origin and are subject to somewhat similar laws.

55. Velocities of Emission of Thermions.—The velocities with which thermions are emitted from the surface of the emitter can be measured in the same way in which the velocities of photoelectrons are measured (Sec. 48). The velocities of the thermions are found to be very much smaller than those of photoelectrons and to be distributed according to Maxwell's law, just as if they were molecules in a gas at the temperature of the emitter.

To see clearly what this latter statement means, suppose that the electrons emerge into a space entirely surrounded by the emitting substance, all at the same temperature T ; *i.e.*, consider emission into a cavity in the emitter. As electrons accumulate in the cavity, they will begin to return to the emitter, and an equilibrium condition will come into existence in which as many electrons return to the emitter per second as issue from it. The cavity is then filled with an *electron gas* in equilibrium at temperature T . The density of this gas is so low that it will obey the laws of perfect gases. The rate at which molecules return to the emitter from such a gas can be calculated from kinetic theory in terms of the density and temperature of the gas. If the assumption is made that all of the electrons enter the emitter as they strike its surface, none of them being reflected, then the rate of return must equal the rate of emission as given by Richardson's

¹ Quoted from HUGHES and DUBRIDGE, *loc. cit.*

equation. In this way, the density and pressure of the electron gas corresponding to any temperature can be calculated. The emitted electrons must be distributed in velocity in the same manner as are the returning electrons. Hence, their distribution in velocity, also, can be calculated, from Maxwell's law for molecular velocities in a gas. It follows from kinetic theory that their mean kinetic energy as they come from the emitter must be $2kT$; this amounts only to 0.22 electron-volt at 1000°C . (An *electron-volt* is the work done on an electron as it passes from one point to another between which there is a potential difference of 1 volt. It equals

$$4.803 \times \frac{10^{-10}}{300} = 1.601 \times 10^{-12} \text{ erg.})$$

As a concluding remark concerning thermionic emission, it may be mentioned that if an accelerating electric field is applied to the emitter, the maximum thermionic current is somewhat increased. As a matter of fact, very intense fields, of the order of 10^8 volts per cm., are able to cause small currents of electrons to flow from metals even at room temperature. These are sometimes called "field currents."

56. Theories of Electrons in Metals.—An explanation of all the features of photoelectric and thermionic emission can be given only on the basis of a theory of the electrical constitution of matter. At this point, therefore, the electron theory of metals will be very briefly discussed.

Soon after the discovery of the electron, it was suggested that electrical conductivity in metals may be due to the existence, in the interatomic spaces, of "free" electrons, which, drifting under the influence of an applied electric field or potential gradient, cause the transport of charge, the actual direction of motion of the electrons being opposite to the usual convention concerning the *positive* direction of current. This theory was developed into the well-known *classical* theory of metallic conduction by Drude, Lorentz,¹ and others. They assumed that the free electrons move about and collide with each other just as do the molecules of a gas, having the usual maxwellian distribution of velocities proper to a gas of electrons at the temperature of the metal. They showed that such free electrons would account, also, for the high *thermal* conductivity of metals; because of their light mass, the electrons would have, according to classical theory, much higher velocities of thermal agitation than do the atoms, and for this reason they would be, in proportion to their numbers, much more effective in the conduction of heat.

¹ LORENTZ, H. A., "Theory of Electrons."

The theory had some success in accounting for the properties of the electrical and thermal conductivities of metals. Most of the formulas that were obtained, however, contained as constants the number of free electrons per unit volume and also the mean free path of the electrons among the atoms, quantities that were not independently known. Furthermore, a serious difficulty was encountered in connection with specific heats. The specific heat of a metal is fully accounted for by the kinetic and potential energy which the atoms themselves should possess, according to the classical principle of the equipartition of energy (see Sec. 85). It was necessary to suppose, therefore, that for some reason the heat energy of the free electrons was very small, or at least independent of temperature. Several lines of evidence indicated, however, that the number of the free electrons should be of the order of magnitude of the number of atoms; and, according to the principle of equipartition, a free electron should have the same average kinetic energy as an atom. Thus, it was hard to understand why the specific heats of good conductors were not considerably larger than the experimental values.

This last difficulty was removed completely in a new theory introduced by Sommerfeld in 1928. His theory may be regarded as a first step toward the modern wave-mechanical theory of metals. The latter theory has been developed to a certain point during the last few years, but it remains incomplete and rather complicated. Its conclusions are reproduced in the Sommerfeld theory with sufficient completeness and accuracy for the application to the more important features of photoelectric and thermionic emission, so the Sommerfeld theory will be used here.

The concept of a gas of free electrons within the metals is retained by Sommerfeld, but this gas is treated in terms of wave mechanics. Because of their light mass and their high density in the metal, the electrons form a degenerate gas of "Fermi-Dirac type," as it is called, after the two physicists who first worked out the theory of such a gas. The electrons have a distribution of velocities very different from the classical one as represented by Maxwell's law.

According to classical theory, the electrons should all have zero energy of translation at absolute zero. As the temperature rises, their energies of translation should increase and at $T^\circ\text{K.}$ should have an average value per electron of $\frac{3}{2}kT$, the distribution of energy among the electrons being the familiar Maxwell distribution curve shown in Fig. 27 for temperatures of 300°K. and 1500°K. A few electrons have high energy, but the majority are grouped around the mean value $\frac{3}{2}kT$, which at 1500°K. is equivalent to some 0.2 electron-volt.

For comparison, the distribution of energy given by the Fermi-Dirac theory is shown in Fig. 28 for the electrons in platinum at three

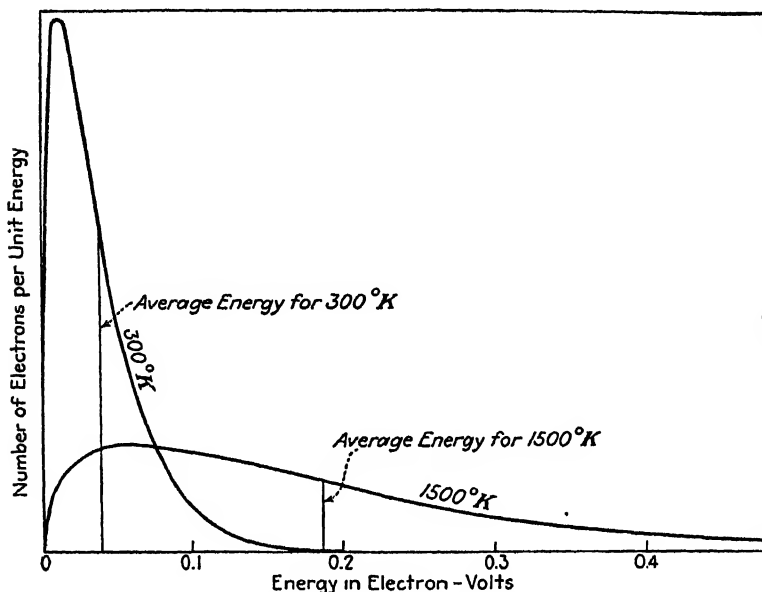


FIG. 27.—Distribution of energy among conduction electrons at temperatures of 300° and 1500°K., according to the classical distribution law of Maxwell.

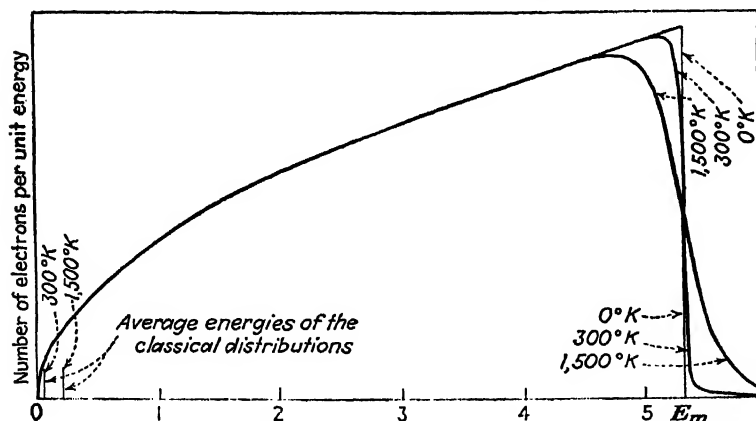


FIG. 28.—Distribution of energy among the conduction electrons of platinum at three different temperatures, according to the Fermi-Dirac-Sommerfeld theory.

temperatures: absolute zero, 300, and 1500°K. It is to be observed that the Fermi-Dirac distribution does not give zero energy for all free electrons at absolute zero, as is required by the classical theory.

Instead, the graph for absolute zero starts from the origin and rises according to an equation of the type $y = a\sqrt{x}$ until a maximum energy E_m is reached, at which point the graph drops suddenly to zero. No electrons possess energy greater than E_m . According to the theory, the value of E_m is given¹ by

$$E_m = \left(\frac{h^2}{2m}\right)\left(\frac{3n}{8\pi}\right)^{2/3},$$

where h is Planck's constant, m is the mass of the electron, and n is the number of electrons per unit volume. If it be assumed that the number of free electrons per unit volume is of the same magnitude as the number of atoms per unit volume, then E_m for platinum (6.6×10^{22} electrons per cubic centimeter) is approximately 6 electron-volts. The theory shows that the mean value \bar{E} of the energy per electron at absolute zero is

$$\bar{E} = \frac{3}{5} E_m = 3.6 \text{ electron-volts for platinum.}$$

This is to be compared with the corresponding mean values given by the classical theory, *e.g.*, 0.2 electron-volt for 1500°K. quoted above.

These results are astonishing to any one who is used to thinking in classical terms, for they mean that even at absolute zero the free electrons in a metal have enormous energies. It is questionable, however, whether it quite makes physical sense to say that this energy is kinetic energy of *motion*. In wave mechanics many ordinary terms lose half of their usual physical meaning when applied to atomic phenomena.

The *rate of increase* of energy with temperature, according to the Fermi-Dirac theory, is much less than it is according to classical theory. Referring again to the distributions in energy for platinum at 300°K. and at 1500°K. as shown in Fig. 28, we note that only in the neighborhood of E_m is there any significant change in distribution with temperature. As the temperature rises, the sharp discontinuity for absolute zero at E_m becomes "rounded off" more and more and an increasing number of electrons come to possess energy greater than E_m . A few electrons even have energies up to several times E_m ; the curve has a long tail (extending, theoretically, up to all energies), and in this tail it can be shown that the distribution of energies follows approximately Maxwell's law, just as in the classical theory of gases.

From the curves it is obvious, furthermore, that the *total energy* of the free electrons changes very little with rise of temperature.

¹ Cf. KENNARD, E. H., "Kinetic Theory of Gases," Secs. 238, 240, 1938.

Hence the free electrons will contribute little to the *change* in energy of the metal as its temperature rises, *i.e.*, to its specific heat. Thus the difficulty noted above in regard to specific heats disappears.

57. Origin of Photoelectrons and Thermions.—Either of the two theories of metals that have been described furnishes at once an explanation of thermionic emission. Presumably, the free electrons are retained in the metal by forces of attraction. There must be as many positively charged atoms in the metal as there are free electrons, and the electrons will be held by the attraction of these positive charges. If, for any reason, an electron escapes from the metal, as it moves away, it will be drawn back by an attraction that may be regarded as arising from its "electrical image" in the surface of the metal, this image being caused by the repulsion of the other electrons by the escaping one and the consequent exposure of positive charge. Thus, in escaping, the electron will lose a certain amount of kinetic energy, for which we may write $e\phi$, e being the numerical electronic charge.

Now in the metal there will always be a few free electrons with energies of thermal agitation exceeding $e\phi$. These electrons will be able to escape as thermions. As the temperature rises, the number of electrons with energies greater than $e\phi$ will increase rapidly; hence, the thermionic current will increase rapidly with rising temperature, as is observed to be the case. If, however, we adopt the Sommerfeld theory, an important question at once arises. If, even at low temperatures, a metal possesses some free electrons with energies as large as E_m , the range of value of which is from 5 or 6 electron-volts up, why do we not have thermionic emission of electrons at all temperatures above absolute zero, since we found that the work function was in general less than 6 volts? The answer is that on the new theory $e\phi$ does not represent the work Ω done by the electron against attractive forces in escaping. The latter work is equal not to $e\phi$ but to $e\phi + E_m$. For the thermions come from the effective top of the Fermi distribution and so start with energy E_m . The additional energy required to remove an electron as a thermion is, therefore, not Ω but only $\Omega - E_m$. This is the quantity that appears as $e\phi$ in the thermionic equation¹

¹ Strictly speaking, $e\phi = \Omega - E_m$ only at 0°K. The question as to a possible variation of the heat of emission of an electron with temperature cannot be discussed here; in any case, the experimental facts themselves are not clear. It might be thought, perhaps, that since only electrons with energies greater than Ω actually escape, no energy at all would need to be supplied in order to remove the thermions. As such high-speed electrons escape, however, a redistribution of velocities of the remaining electrons occurs, principally at the expense of electrons with energies near E_m , since it is only electrons with energies near to or greater than E_m whose

[(54) in Sec. 53], so that in that equation

$$e\phi = \Omega - E_m. \quad (56)$$

For platinum ϕ is 6.3 volts and E_m/e is at least 6 volts; hence Ω/e is at least 12.3 volts. For the metals in general, Ω/e is 5 to 20 volts higher than the ordinary "work function" ϕ . These high values of Ω are qualitatively confirmed by observations on the index of refraction of electron waves (*cf.* Sec. 117 in Chap. VII).¹

The relation between these various quantities is shown diagrammatically in Fig. 29. The heavy curve shows the way in which the *potential* energy of an electron varies as the electron, coming from

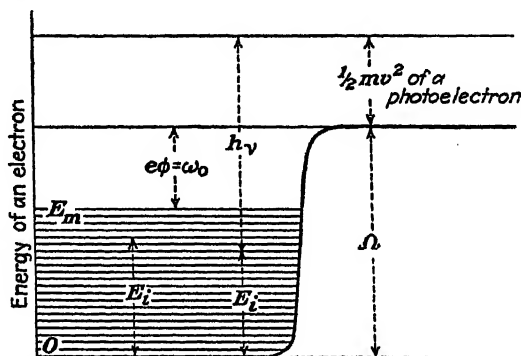


FIG. 29.—Energy of electrons in a metal. An electron with extra energy $e\phi$ can just escape as a thermion. A photoelectron absorbs $h\nu$ and retains $\frac{1}{2}mv^2$ as kinetic energy.

the left in the figure, passes through the surface of the metal into the space outside, the potential energy rising by the amount Ω . The ordinate of any point above this line may represent the total energy of an electron; the height of the point above the heavy line will then represent its kinetic energy. The horizontal lines at the level marked E_m and below indicate the Fermi energy levels (actually much more closely spaced) at 0°K.; at any other temperature, there are a few electrons with energy above E_m .

Concerning *photoelectrons*, on the other hand, there was long a difference of opinion as to whether they come from among the free electrons or out of the interior of the atoms. On either hypothesis, it is easy to explain the fact that all velocities occur up to a maximum value v_m given by the Einstein equation

$$\frac{1}{2}mv_m^2 = h\nu - \omega_0.$$

distribution is subject to change. Thus the result is the same as if the thermions themselves started with energy E_m .

¹ See also THOMSON, G. P., "The Wave Mechanics of Free Electrons," 1930.

The quantity ω_0 is interpreted here as the energy lost by an electron in escaping through the surface of the emitter. But light is known to penetrate a distance equal to many atomic diameters into the metal; hence, photoelectrons should originate at various depths. In order to reach the surface, the electron would have to pass by (or through) many atoms, and collisions would be probable, at each of which the electron might lose some of the energy $h\nu$ initially given to it by the light. The electron would reach the surface with less energy than $h\nu$, would give up energy ω_0 in escaping from the surface, and would emerge with energy less than the maximum. Some electrons would lose almost all of their initial energy in this way and would emerge with velocity almost zero.

If, now, the photoelectrons come from among the free electrons, they will have an initial energy E_i due to thermal agitation and, after absorbing $h\nu$ from the light, will start off with energy $h\nu + E_i$. The photoelectrons produced near the surface should thus emerge with energy

$$\frac{1}{2}mv^2 = h\nu + E_i - \omega_0.$$

Since E_i has no upper limit, there should exist, theoretically, no maximum velocity for a given frequency ν , and no sharp threshold ν_0 , the curves tailing off instead of plunging sharply into the axis. According to classical theory, however, the value of E_i is small; according to the classical principle of the equipartition of energy, or Eq. (104) in Sec. 84, the average value of E_i is

$$\bar{E}_i = \frac{3}{2}kT,$$

where T is the absolute temperature and k is the Boltzmann constant (1.381×10^{-16} erg deg.⁻¹). At 300°K., \bar{E}_i is equivalent to about 0.04 electron-volt; at 1500°K., to about 0.20 electron-volt. Practically, therefore, a fairly definite maximum velocity v_m and threshold frequency ν_0 would exist. But, since \bar{E}_i is proportional to T , the apparent values of v_m and ν_0 should vary somewhat with temperature. Most of the experimental evidence indicates that a variation of the right order of magnitude does not exist, at least not in all cases. Formerly, this was commonly accepted as evidence that the photoelectrons come from the atom and not from the free electrons, since there was no reason for believing that the *internal* energy of atoms depends on temperature over the range covered by the experiments.

This latter view was strengthened by observations on the photoelectric effect with X-rays, the frequencies of which are several thousand times those of ordinary light. In these experiments (see Sec. 196)

there is very direct evidence not only that photoelectrons come from atoms but that they come from different "depths" within atoms and, in order to leave the atom, must expend quantities of energy which can be directly measured and which can be checked by independent means.

Further, the discovery of the Compton effect with X-rays (see Sec. 198) and its explanation emphasized the importance of another objection to the theory that the photoelectrons come from the free electrons. It would be impossible for a free electron to absorb a quantum of radiation, for the laws of the conservation of energy and of momentum could not simultaneously hold. If an electron, free and at rest, acquires kinetic energy $\frac{1}{2}mv^2$ by the absorption of a quantum $h\nu$, we should have, from the laws of conservation of energy and of momentum,¹

$$h\nu = \frac{1}{2}mv^2, \quad \frac{h\nu}{c} = mv,$$

where c is the velocity of light, m the mass of the electron, and v its velocity after absorbing the quantum. These equations can hold simultaneously only if $v = 2c$ —an impossibility. But if the electron absorbs light while in an atom, the momentum of the rest of the atom may change, and the above laws of mechanics may hold simultaneously.

If, however, the photoelectron originates in the atom,² a new difficulty arises. The electron will then have to do work in order to escape from the atom into the interatomic space, and this work will be included in ω_0 . Thus, ω_0 would necessarily exceed the work of escape $e\phi$ of a thermion, which starts presumably in the interatomic space. The difference should, in fact, be relatively large, for it ought to be of the order of the work required to ionize an atom, which, at least for atoms in the vapor state, amounts always to several electron-volts. Yet it is found experimentally that ω_0 and $e\phi$ differ (at least usually) by less than 1 electron-volt, often by less than 0.1 electron-volt.

Thus, the classical theory was not able to furnish a satisfactory answer to the question as to the origin of the photoelectrons. The new theory, on the other hand, removes all objections to supposing that photoelectrons liberated by visible or ultraviolet light (in contrast with most of those liberated by X-rays) come from the so-called "free" electrons. For, according to the new theory, these electrons are not really free; they are simply bound to the metal as a whole

¹ See Sec. 198(a) for the momentum of a photon.

² We are considering photoelectrons ejected by visible and ultraviolet light, not by X-rays.

rather than to an individual atom. They differ from the electrons in the interior of an atom chiefly in that they are subject to forces exerted by several atoms at once. The difficulty regarding conservation of momentum, therefore, does not arise.

The situation is improved, also, in regard to the variation with temperature. The electrons that emerge with maximum kinetic energies will be chiefly those that come from the top of the Fermi distribution (E_m in Fig. 29). After absorbing a quantum $h\nu$, one of these electrons has energy $h\nu + E_m$. Then it must do work Ω to escape from the surface, so that, if no further loss of energy occurs, it emerges with kinetic energy

$$\frac{1}{2}mv^2 = h\nu + E_m - \Omega. \quad (57)$$

This value of $\frac{1}{2}mv^2$ will represent the maximum value of the kinetic energy of the photoelectrons. The equation has the same form as the Einstein equation [(53) in Sec. 49], which was

$$\frac{1}{2}mv_m^2 = h\nu - \omega_0.$$

Thus, the ω_0 of the Einstein equation should have the magnitude

$$\omega_0 = \Omega - E_m. \quad (58)$$

Comparing this equation with Eq. (56) or

$$e\varphi = \Omega - E_m,$$

where φ is the thermionic work function, we see that ω_0 should also equal $e\varphi$, in agreement with experiment.

Even on the Sommerfeld theory, however, there ought, theoretically, to be no true limit to the maximum energy of ejection, and hence no frequency threshold; for a *few* free electrons are present with *all* energies. The curves in Figs. 23 and 24 ought, therefore, to be rounded off at the end, so as to meet the axis asymptotically instead of plunging into it at a finite angle (except at $T = 0$). This effect would be much smaller than the effect required by classical theory, however, for the number of electrons in the maxwellian tail of the Fermi-Dirac distribution is extremely small. Furthermore, there is experimental evidence that a slight rounding off of the curves at this point does exist.¹

¹ DuBRIDGE, *Phys. Rev.*, vol. 43, p. 727 (1933).

CHAPTER IV

THE THEORY OF RELATIVITY

The greatest revolution in physical thought during the twentieth century undoubtedly has been that which resulted from the introduction of the quantum theory. This, however, was not the first innovation in modern physical theory to attract general attention. In a different way, the new form of relativity put forward by Einstein in 1905 made an equally complete break with classical ways of thought. Relativity touches all branches of atomic physics here and there, but in itself it occupies a detached position, more fundamental than that of other parts of physical theory. For this reason it seems appropriate at this point to devote a chapter to a discussion of the theory of relativity. Any student who does not wish to interrupt the study of atomic physics at the present juncture can omit the present chapter without appreciable embarrassment in reading subsequent chapters.

58. Newtonian Relativity.—The phenomenon of *motion* has been a subject of speculation since ancient times. It was early recognized that all motion involves displacement *relative* to something or other; but ideas have varied in regard to the entity relative to which the displacement occurs. In his treatise on mechanics, Newton says that "absolute motion is the translation of a body from one absolute place to another absolute place."¹ But he does not explain what he means by "absolute place." He states explicitly that the physicist can detect translatory motion only in the form of motion *relative to other material bodies*.

Motion involves, also, the passage of time. Until recently, time was regarded as something essentially distinct from space or from the behavior of material bodies. Newton says, "Absolute, true, and mathematical time, of itself, and by its own nature, flows uniformly on, without regard to anything external."² Thus there was supposed to be a single time scale valid everywhere. Until 1905, this view seems to have been accepted universally.

The kind of relativity embodied in these views has been called "Newtonian relativity." For purposes of comparison, it will be worth while to formulate it in mathematical terms.

¹ Cf. MACH, E., "Science of Mechanics," 4th ed., p. 226, 1919.

² *Ibid.*, p. 222.

Let us represent position by means of Cartesian coordinates x, y, z and let t denote the time. Then a set of four numbers representing values of x, y, z , and t specifies the position and the time at which an *event* of some sort occurs. The event might, for instance, consist in a projectile's passing a certain point in space at a certain time, or it might consist in an electron's leaving a filament or entering a Geiger counter. All physical phenomena, in so far as they involve positions and times, can be resolved into sequences of events.

In order that given values of our x, y, z may fix the location of an event, we must have some material reference body from which distances can be measured, such as the floor and walls of a laboratory, or the bench marks established by surveyors on the earth. Similarly, in order to define a time, we must have available some reference process, such as the rotation of the earth or the swinging of a pendulum, in order that times may be specified by stating the stage to which the reference process has advanced. These material means of fixing positions and times, together with the methods adopted for using them, are said to constitute a *space-time frame of reference*.

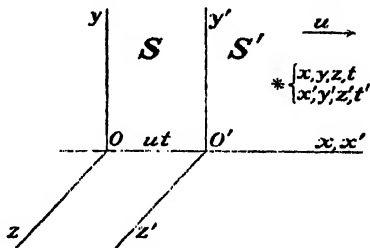


FIG. 30.--Two frames of reference in uniform relative translatory motion.

Now suppose we have *two different* frames of reference, each in *uniform translatory motion* relative to the other. One frame might be on the earth, while the other is located on an airplane flying overhead. Let us call the two frames S and S' ; and let the velocity of S' relative to S be u . Let coordinates and times of any event as obtained when the frame S is used be denoted by x, y, z, t ; and let those obtained for the same event when S' is used be denoted by x', y', z', t' . In order to make the relation between these variables as simple as possible, let us choose our axes so that the x - and x' -axes are both parallel to the velocity u and, in fact, so that these axes slide along each other; and let the y' - and z' -axes be parallel to y and z , respectively (Fig. 30). Let us also agree to count time from the instant at which the two origins of coordinates, O and O' , momentarily coincide. Then the coordinates of O' , the origin of S' as measured in the first frame S , will be $x = ut, y = 0, z = 0$; and, if any event occurs at a position and time specified by certain values of x, y, z, t , then, according to Newton, its position and time as measured using S' will be represented by x', y', z', t' , where

$$\begin{aligned}x' &= x - ut, & t' &= t, \\y' &= y, \\z' &= z.\end{aligned}$$

These equations may be called the *equations of transformation* for Newtonian relativity. They enable us to pass from coordinates and times of events referred to one frame of reference over to coordinates and times referred to another, when all relations are assumed to be Newtonian. The equation, $t' = t$, which may seem superfluous but is included with an eye to the future, expresses the fact that in the Newtonian theory there is supposed to be a single time scale valid for all frames of reference.

59. Relativity and the Propagation of Light.—With the adoption of the wave theory of light, a new element, unknown to Newton, was brought into the problem of motion. For, if light consists of “waves” in an ether, these waves should move with a definite velocity *relative to the ether*, and their velocity *relative to material bodies* should change when the motion of these bodies through the ether changes. Analogous statements made about waves in material mediums are certainly true. Water waves of a given length, for example, move with a certain speed over the water; as seen from a moving ship, they move past the ship faster when the ship is moving over the water against the waves than when the ship is moving in the same direction as the waves themselves. Now, in accordance with Huygens’ principle, many optical phenomena depend in a very simple way upon the velocity of light relative to material objects. It might be anticipated, therefore, that such phenomena would be influenced by a motion of the optical apparatus through the ether. The velocities

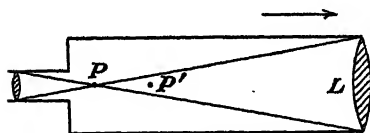


FIG. 31.—Effect of motion of a telescope through the ether.

that can be given to a material body in the laboratory are extremely small as compared with the velocity of light; but this is less true of the velocity of the earth in its orbital motion about the sun, which is one ten-thousandth of the velocity of

light in free space. An influence of the earth’s orbital motion upon optical phenomena might therefore be anticipated.

An interesting case to consider is the formation of images by the object lens of a telescope. Suppose a light wave from a star enters the telescope sketched in Fig. 31; when the telescope is at rest in the ether, let the wave come to a focus in such a way as to form a star image on the cross hairs at P . Then, if the telescope, instead of being

at rest, is moving toward the star, the wave might be expected to focus on the same point P in the ether as before. While the wave is passing from the lens to this point, however, the telescope will move forward, carrying the cross hair to some other point P' . The eyepiece would therefore have to be drawn out farther in order to focus on the image of the star. Thus the focal length of the telescope would appear to the observer to be increased. Similarly, if the telescope were moving in the same direction as the light, its effective focal length would be shortened. If it were moving at right angles to its axis, on the other hand, there would be only an apparent lateral displacement of the star, constituting the familiar phenomenon of aberration. All of these effects can also be thought of as due to motion of the ether in the opposite direction relative to the telescope, or to what has been called an "ether wind."

As the telescope is pointed at stars in different directions, the apparent focal length of any astronomical telescope might therefore be expected to vary slightly because of the earth's orbital motion. But *no such effect has ever been observed.*

Fresnel long ago pointed out a way in which it might happen that effects of this sort *do not occur*. It might be that moving transparent bodies, such as a lens, partially drag the light waves along with them. In the case just considered, if the lens L were to drag the light with it in its motion (toward the right in Fig. 31), the part of the wave that goes through the center of the lens would spend a longer time in the lens, and hence would be retarded more than it would be if there were no motion. Consequently the wave would emerge from the lens more concave in shape, and would focus on a point nearer the lens. If the amount of the drag were just right, there might even be no effect at all on the apparent focal length of the telescope, the star image falling on the cross hair however the telescope might be moving. Fresnel showed¹ that all such effects on phenomena of refraction would be prevented if it were a law of optics that any moving transparent medium of refractive index n changes the velocity of light waves in such a way as to add vectorially to their velocity in the stationary medium, which is c/n , the fraction

$$1 - \frac{1}{n^2}$$

of the velocity of the medium. Interestingly enough, Fizeau showed experimentally in 1851, by an interference method, that a moving

¹ For an excellent discussion of relativity see M. Born, "Einstein's Theory of Relativity," translated by H. L. Brose, Methuen & Co., Ltd., London, 1924.

column of water *does* drag the light waves with it to the exact extent required by Fresnel's formula!

The cause of this effect on the light waves was supposed by Fresnel to lie in an actual partial dragging of the ether itself along with the moving medium, analogous to the dragging of the water by a fish net as it is drawn along. When Lorentz worked out his electromagnetic theory, however, about 1895, on the assumption that the ether remains always at rest, he found that the theory led automatically to the Fresnel formula for the drag of the light waves by moving matter.

If the Fresnel drag of the light waves is assumed, it can be shown that there can never be any effect on ordinary optical phenomena that is of the *first order* in the velocity of the apparatus through the ether. There might, however, be second-order effects. Since the square of the earth's orbital velocity is only a hundred-millionth of the square of the velocity of light, it might be thought that such effects would be far too small to detect. In seeking for a sufficiently delicate means of observation for this purpose, Michelson was led to invent his interferometer. With this instrument, in conjunction with Morley, he performed in 1887 a famous experiment in which a second-order

effect could have been detected if it had occurred.

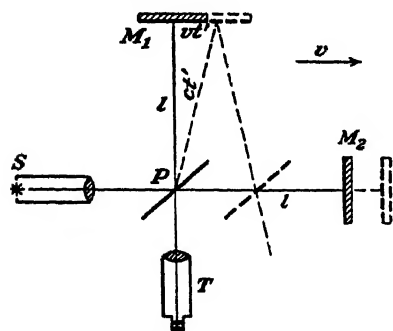


FIG. 32.—Diagram illustrating the Michelson-Morley experiment.

60. The Michelson-Morley Experiment.—

The interferometer arrangement that was used in this experiment is sketched in Fig. 32. A beam of light from a lamp *S* falls upon a half-silvered glass plate *P* placed at 45° to the beam, which divides each wave into two parts. One wave, reflected from *P*, travels off sideways to a mirror *M*₁, which reflects it back to *P*; part of it is then transmitted through *P* and enters the telescope *T*. The other wave, transmitted at once through *P*, travels ahead and is reflected back by a second mirror *M*₂; upon returning to *P*, it is partially reflected there into the telescope on top of the first wave, with which it forms an interference pattern.

Let both mirrors be at the same distance from the plate *P*. Then, if the apparatus is at rest in the ether, the two waves take the same time to return to *P* and meet in phase both there and in the telescope. Suppose, however, that the apparatus is moving with speed *v* through the ether in the direction of the incident beam of light. Then, if the

incident wave strikes the plate P when it has the position shown in the figure, the paths of the waves and the subsequent positions of reflection from mirrors and plate will be as shown by the dotted lines. The necessary change in the direction of the transverse beams is produced automatically, as an aberration effect, through the operation of Huygens' principle. But now the times taken by the two waves on their journeys are no longer equal. The wave moving longitudinally toward M_2 has a velocity *relative to the apparatus* of $c - v$ on the outgoing trip, c being the speed of light through the ether, and $c + v$ on the return trip. Hence the time taken by this wave to get back to the plate is

$$t_2 = \frac{l}{c - v} + \frac{l}{c + v} = \frac{2cl}{c^2 - v^2},$$

l being the distance from the plate to either mirror. If we keep small terms only as far as v^2/c^2 , we can write, by the binomial theorem,

$$\frac{1}{c^2 - v^2} = (c^2 - v^2)^{-1} = c^{-2} + v^2 c^{-4}.$$

Hence

$$t_2 = 2l \left(\frac{1}{c} + \frac{v^2}{c^3} \right).$$

The wave moving transversely, on the other hand, travels along the hypotenuse of a right-angled triangle, one side of which is of length l , and therefore it has farther to go. Let it take a time t' to go from P to M_1 , traveling a distance ct' through the ether. Then in the same time the mirror M_1 advances a distance vt' . Hence

$$c^2 t'^2 = v^2 t'^2 + l^2, \quad t' = \frac{l}{(c^2 - v^2)^{1/2}}.$$

But, using the binomial theorem again,

$$(c^2 - v^2)^{-1/2} = c^{-1} + \frac{v^2 c^{-3}}{2},$$

as far as terms in v^2/c^2 . Hence this wave returns to P after a time

$$t_1 = 2t' = 2l \left(\frac{1}{c} + \frac{1}{2} \frac{v^2}{c^3} \right).$$

Thus the two waves interfere in the telescope with a phase difference of $t_2 - t_1 = lv^2/c^3$ seconds or $c(t_2 - t_1)/\lambda = lv^2/(c^2\lambda)$ wave lengths, λ

being the wave length of the light. The fringe pattern is accordingly shifted by the motion through

$$\frac{v^2 l}{c^2 \lambda}$$

fringes.¹

In performing the experiment, the whole apparatus, floated on mercury, was rotated repeatedly through 90°. Since the two light paths are caused to exchange roles by such a rotation, it should cause the fringe pattern to shift by twice the amount just calculated or by

$$N = 2 \frac{v^2 l}{c^2 \lambda}$$

fringe widths. By reflecting the beam back and forth several times, the effective length l was brought up to 11 meters. Even then, with a wave length of about 5.9×10^{-5} cm., if we insert for v the whole orbital velocity of the earth so that $v/c = 10^{-4}$, we find only

$$N = 2 \times 10^{-8} \times \frac{11 \times 10^2}{5.9 \times 10^{-5}} = 0.37,$$

or about one-third of a fringe. Michelson and Morley were sure, however, that they could detect a shift of a hundredth of a fringe. *Such shifts as were observed amounted only to a small fraction of the theoretical value and were not consistent.* Thus the result of the experiment was negative; the expected ether wind did not appear.

It might happen, to be sure, that at a given moment the earth by accident had no resultant component of velocity parallel to the surface of the earth; for upon its orbital motion there would probably be superposed a motion of the entire solar system through the ether. But, if this happened at a certain time, then 6 months later the earth's orbital velocity about the sun would be reversed, and its velocity through the ether should then be twice its orbital velocity. Michelson and Morley made observations not only at various times of day but also at different seasons of the year, always with the same negative result.²

From the theoretical point of view, Michelson and Morley's failure to detect the anticipated motion of the earth through the ether was one of the most important experimental results ever obtained.

¹ For the formation of the fringes in the interferometer see C. R. Mann, "Manual of Advanced Optics."

² MICHELSON and MORLEY, *Silliman J.*, vol. 34, pp. 333, 427 (1887); MORLEY and MICHELSON, *Phil. Mag.*, vol. 24, p. 449 (1887). Cf. MILLER, *Rev. Modern Physics*, vol. 5, p. 203 (1933).

It was very hard to bring into harmony with current theories of light and matter. The theoretical calculation rests on a peculiarly simple foundation, for the only properties of light that are made use of are the constancy of its velocity in space and Huygens' principle. Neither of these principles can be given up if we are to retain the idea of waves in a mechanical ether. Only three possible explanations of the negative result seem to offer themselves.

1. Perhaps the student will have wondered why we cannot simply assume that the earth drags the ether with it, much as a moving baseball carries along the air next to it. On this assumption there would never be any motion of the earth through the ether at all, and no difficulties could arise. The objection to this explanation is that the ether next to the earth would then be in motion relative to the ether farther away; and this relative motion between different parts of the ether would cause a deflection of the light rays coming from the stars, just as wind is observed to deflect sound waves. This deflection would alter the amount of the stellar aberration.¹ That is, the ether wind would then occur outside of the earth. It has been found very difficult to devise a plausible type of motion for the ether which would give a value of the aberration agreeing with observation and yet at the same time preserve the negative result of the Michelson-Morley experiment. Furthermore, doubt is thrown upon such an explanation by the fact that, as has already been stated, experiments show that a transparent object of laboratory size does not drag the light waves with the full velocity of the moving matter, as it necessarily would do if it *completely* dragged the ether along with it; and the observed *partial* drag is fully accounted for by current electromagnetic theory, which is based on a stationary ether.

2. As an alternative, the negative result would at once be accounted for if we could assume that light projected from a moving source has added to its own natural velocity the velocity of the source, just as the velocity of a projectile fired from a moving ship is equal to the vector sum of its velocity of projection from the gun and the velocity of the ship. In a vacuum, light would then leave its source always with the same velocity relative to the source. If this were true, the negative result obtained by Michelson and Morley would at once be explained, for the light from their lamp would have always the same constant velocity in all directions relative to the lamp and to the interferometer. Such an assumption, however, is in gross conflict with the wave theory of light. It is of the essence of waves that they have a definite velocity

¹ Stellar aberration is the apparent change in direction of a star due to its orbital motion. See any text on astronomy.

relative to the medium which transmits them, just as sound waves have a definite velocity relative to the air. Furthermore, there is strong *experimental evidence against* supposing that the velocity of light is influenced by a motion of the source from which it comes. This evidence is furnished by several experiments in the laboratory, and by various astronomical observations, which we have not space to describe. **1**

3. The third possible explanation of Michelson and Morley's negative result is one that was put forward by Fitzgerald and by Lorentz working independently. They suggested that motion through the ether might, in some manner not yet understood, cause the material of which the interferometer was constructed to *shorten in a direction parallel to the motion*. It is easily seen that such a contraction in the ratio $\sqrt{1 - v^2/c^2}$ would serve to equalize the light paths and thus to prevent a displacement of the fringes. Lorentz endeavored to make the occurrence of such a contraction plausible on the basis of his electrical theory of matter, but he was never quite able to show that it *must* occur. This explanation came to be the favored one, but it was never felt to be very satisfactory because of its *ad hoc*¹ nature.

It might be remarked finally that nothing is gained by carefully avoiding all reference to a physical ether and speaking only of velocities relative to some frame of reference. For example, we might use a frame of reference fixed in the sun; then, to explain astronomical aberration, we should need to assume that light moves with a constant velocity relative to this frame, and on this assumption the difficulty about the Michelson-Morley experiment would arise as described above.

61. The New Relativity of Einstein.—From the situation just described one easily gains the impression that there exists something like a conspiracy in nature to prevent us from detecting motion through the ether. A similar situation can be shown to exist in the field of electromagnetism as well as in optics. A number of electrical or magnetic experiments can be invented which, at first sight, offer promise of revealing motion through the ether; but always there occurs some other effect which just cancels the effect sought. Sometimes it is actually the Lorentz-Fitzgerald contraction just described or the forces associated with this contraction by which the cancellation is effected.

¹ An *ad hoc* hypothesis is one that is proposed in order to explain a single fact as distinguished from a hypothesis that explains simultaneously several distinct facts. "*Ad hoc*" translated from the Latin means "to this" or "for this."

In reflecting upon this extraordinary situation, Einstein arrived in 1905 at a radically new view of the whole subject.¹ He proposed the view that, for some reason not yet known, motion through the ether is a *meaningless concept*; only motion *relative to material bodies* has physical significance. He then set about investigating how this assumption could be made to harmonize with the known laws of optics. Possibilities of conflict arise in any argument involving the velocity of light. Suppose, for example, a frame of reference S' (say on the earth) moving relative to another frame S (say on the sun) carries a source of light. Then light from this source must move with the same velocity relative to S as light from a source on S itself, since, as explained above, we cannot suppose that the velocity of light is influenced by motion of its source. But this light must also move with the same velocity *relative to S'* ; otherwise the laws of nature would not be the same on S' as on S , and no reason could be assigned for their being different. Yet it seems quite impossible that light should move with the same velocity relative to *each of two frames that are moving relatively to each other!*

Einstein accordingly put the laws of the propagation of light in the forefront of the discussion. He presented his new theory as based upon two propositions, which are known as the postulates of the *special or restricted theory of relativity*. They may be stated as follows:

1. *The laws of physical phenomena are the same when stated in terms of either of two inertial frames of reference (and involve no reference to motion through an ether).*
2. *The velocity of light is independent of the motion of its source.*

By an "inertial" frame is meant one in which the law of inertia is true, *i.e.*, relative to which a body that is free from external influence and at rest remains at rest. An inertial frame is, therefore, what is usually called an "unaccelerated" one; and the relative motion of two inertial frames can only consist of a uniform translation. Thus, in the special theory of relativity, we compare statements of physical laws expressed in terms of two frames of reference each of which is in *uniform translation* relative to the other.

Of these two "postulates," the second is believed to represent a rather simple experimental fact, whereas the first is a generalization from a wide range of physical experience. There is no implication that the first postulate, which contains the new principle of relativity,

¹ EINSTEIN, *Ann. d. Physik*, vol. 17, p. 891 (1905).

is in any way self-evident; like the assumptions made in all physical theories, it is intended as a hypothesis to be tested by comparing deductions from it with experimental observation.

62. Simultaneity and Time Order.—Einstein found the key by which these two postulates could be brought into harmony in a modification of our ideas concerning time. He showed that it is necessary to give up the Newtonian conception of absolute time. Newton had undoubtedly supposed that it is always possible to say which of two events precedes the other, or that they occur simultaneously, even when they occur at different places. But suppose that the difference in time between two events is less than the time taken for light to travel from the location of one event to the location of the other. Then how could we tell experimentally which event precedes the other?

To determine the order in time of such events, we might place a clock at the location of each one. But how could we set the two clocks into synchronism? We might perhaps attempt to do this, after the old-fashioned practice of navigators, by carrying a chronometer from one clock to the other and comparing the clock readings with the chronometer. But how can we know that our chronometer runs at the same rate while moving as it does at rest? Another method might be to follow the modern practice and exchange radio signals. But then we should need to be able to correct for the time taken by the signal in getting from the location of one clock to the location of the other; and to make this correction requires a knowledge of the velocity of light in *one direction*. Ordinary measurements of the velocity of light furnish only its *average velocity in two opposite directions*. The velocity could be measured in one direction in the same manner as a race is timed, provided we had our clocks already synchronized. This idea results in a logical circle, however. Every other method of measuring the velocity of light in one direction turns out similarly to involve some assumption that cannot be tested in advance. Finally, one might think of hurling a ball from one location to the other with indefinitely great speed, so that no correction for its time of flight would be necessary. If this could be accomplished, it would, in fact, do the trick. But, if all masses increase with velocity as the mass of the electron is known to do, a ball could not possibly be projected with a speed exceeding that of light; therefore this method, too, would fail.

Thus the simultaneity of events and, in part, the time order in general of events at different places, are relative and not absolute concepts. At least this is true so far as physics is concerned. Using a given frame of reference, we can adopt some arbitrary criterion as to the time order

of events that occur in different places; and then, on the basis of this arbitrary criterion, we can set up a system of measurement of space and time. But it is easily shown that, if the same criterion is used for all frames, it is possible for two events to occur at different places but at the same time, as judged in terms of one frame, whereas in terms of another frame they occur at the same place but at different times. The distinction between relations of space and those of time is thus in part a *relative* matter depending upon which frame of reference is used, in somewhat the same way as relations of "right" and "left" depend upon the position of the observer. For this reason it has become customary to speak of space and time as aspects of a four-dimensional continuum known as "space-time."

This fascinating topic, however, cannot be pursued further here. Our main concern must be to derive the formulas that are needed for the application of relativity to physical laws. The general discussion will be closed with the description of a figure that illustrates in a simple way the new theory of space-time relations.

Restricting the discussion to events that happen on the x -axis, let us plot values of x and t as determined by an observer at an instant t_0 when he himself is at P (Fig. 33). Through P we can draw two straight lines representing the progress of light signals that pass the observer just when he is at P , one signal moving toward $+x$ and the other toward $-x$; *i.e.*, any point on one of these lines represents the passage of one of these signals past a certain position on the x -axis at a certain time t . Then the area above these two lines represents events that lie in the *future* for the observer. These events have values of t so greatly exceeding t_0 that the observer still has an opportunity to influence them causally. Below the two lines lies the observer's *past*, consisting of events which may have influenced him causally, and of which he may already have acquired knowledge at the instant t_0 by means of light signals or other means of communication.

There is also a *third* region, however, *between* the lines. This region constitutes the observer's *physical present*. He cannot influence events in this region—there would not be time to get any signal to the location of one of them before it occurred; nor, for the converse reason, can he have any knowledge of them at the time t_0 . *At each point of space these events belonging to the observer's present extend over a finite interval of time*, such as that represented by the line QR in

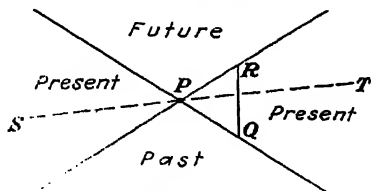


FIG. 33.—Diagram illustrating the classification of space-time relations.

the figure, the interval increasing with distance from P . For an observer in New York this interval amounts to $\frac{1}{60}$ of a second for events in San Francisco; within this $\frac{1}{60}$ sec. (the time required for light to make the double journey), he cannot pick out any unique instant in San Francisco which is simultaneous with the instant t_0 in New York. Here we have an essential difference between Einstein's relativity and Newton's; for the latter reduced the physical present to a single line, such as ST . The observer himself, of course, might set up a system of space-time measurements, in which only the events lying along some such line as ST would be regarded as simultaneous with P . But the choice of this line and the restriction of simultaneity to events on such a line would be arbitrary, depending upon the frame of reference used. All events in the area labeled "present" are simultaneous as far as the observer at P can tell by means of physical observation.

63. The Lorentz Transformation.—If the postulates of Einstein's relativity are accepted, it becomes important to inquire whether or not the accepted laws of physics are all in harmony with the theory. Furthermore, on the basis of this theory, it may be possible to predict new phenomena, which can be looked for experimentally. In any case, the theory must stand or fall according to whether or not the deductions that can be drawn from it agree with the experimental facts.

In order to make deductions from relativity, we must compare the descriptions of some phenomenon in terms of two inertial frames which are moving relatively to each other. It is important to know, first of all, how measurements of space and time compare. For the sake of vividness, we shall often speak of an *observer* who is supposed to make the measurements referred to a particular frame. This must not be understood to imply, however, that relativity has any closer connection than the rest of physics with human psychology.

The ambiguity in regard to times at different places, which was discussed in the last section, is commonly eliminated, in using any given frame of reference, by setting up the time scales in such a way that the velocity of light in vacuum measures the same in any one direction as it does in the opposite direction. The velocity of light in vacuum then becomes the same in all directions and therefore is a universal constant. This adjustment of the time scales can be supposed to be effected experimentally, as it actually is in common practice, by sending light signals back and forth and correcting for the time of propagation on the assumption of a constant velocity of light. Let us use the same units with all frames of reference; to make sure of doing this, we

may select such units of length and time that certain natural constants have the same values for both frames. For example, we may make the wave lengths of spectral lines and the velocity of light measure the same in all frames. When this has been done, the equations expressing all physical laws must take the same form, according to the first postulate of relativity.

Now let S and S' denote two inertial frames, of which S' has a uniform translational velocity u relative to S . Let axes be chosen as before so that the x -axes and the xy planes in the two frames slide along each other (*cf.* Fig. 30); and, at the origins, let us count time from the instant at which these points momentarily coincide. Let t be the time at which some event happens at a point x, y, z , as measured by the observer on frame S , and let x', y', z', t' be the coordinates and time of the same event as measured by the observer on S' . Then it can be shown that the postulates of relativity require that, c being the velocity of light in vacuum,

$$\begin{aligned} x' &= \gamma(x - ut) & y' &= y & \gamma &= \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} & (59) \\ t' &= \gamma\left(t - \frac{ux}{c^2}\right) & z' &= z \end{aligned}$$

These equations, known as the "Lorentz transformation," were discovered by Lorentz in the course of his study of moving matter in an electromagnetic field; but he regarded one frame as at rest in the ether and attached an immediate physical meaning only to measurements made with this frame. In the new theory, on the other hand, all inertial frames are on an equal footing. The equations can be derived from the requirement that the velocity of light shall be the same in both frames together with considerations as to the isotropy of space; the deduction will not be given here,¹ but a proof will be given in Sec. 65 that the Lorentz equations actually do preserve the constancy of the velocity of light.

By solving Eqs. (59) for x, y, z, t , we obtain the inverse transformation:

$$\begin{aligned} x &= \gamma(x' + ut') & y &= y' & (60) \\ t &= \gamma\left(t' + \frac{ux'}{c^2}\right) & z &= z'. \end{aligned}$$

By means of these equations of transformation, positions and times measured in one frame of reference can at once be translated into

¹ *Cf.* SILBERSTEIN, L., "Theory of Relativity," 1924.

positions and times of the same events as measured in the other frame. From the equations, we can easily confirm the statement, made previously, that, according to the new relativity, events which happen at the same place at different times, as viewed from one frame, may be seen from another frame to happen at different places as well. Similarly, a difference in place only, with respect to one frame, may correspond to a difference in both space and time with respect to another. Thus a space difference can be converted partly into a time difference, or vice versa, merely by changing the frame of reference that is being used.

64. Contractions in Space and Time.—There are two cases in which the Lorentz transformation leads immediately to results of special interest.

Suppose a body, when at rest in S , has a length L_0 in the direction of the x -axis. Let it be set moving relative to S at such speed that it is at rest in S' . Its length will then be L_0 as measured in S' ; for its length is determined by natural laws and hence must have a certain fixed value in any frame in which the body is at rest. Let us see how the length now measures in S , relative to which the body is moving with speed u . Before doing this, we must consider what we mean by the *length* of a *moving* object. It seems reasonable to define the length as the distance between two points fixed in S which are occupied by the ends of the object simultaneously, *i.e.*, at the same time t . If the coordinates of these points are x_1 and x_2 , the length is then

$$L = x_2 - x_1.$$

By the same definition, since the body is at rest in S' , its ends have fixed coordinates x'_1, x'_2 such that

$$L_0 = x'_2 - x'_1.$$

If we substitute in this last equation values of x'_2 and x'_1 calculated from the first one of Eqs. (59) for a given value of t , we obtain

$$L_0 = \gamma(x_2 - x_1).$$

Therefore

$$L = \frac{1}{\gamma} L_0 = L_0 \sqrt{1 - \frac{u^2}{c^2}}. \quad (61)$$

From this equation we may draw two distinct conclusions. In the first place, any body measures shorter in terms of a frame relative to which it is moving with speed u than it does as measured in a frame relative to which it is at rest, the ratio of shortening being $\sqrt{1 - u^2/c^2}$. This is a relation between *measurements*.

In the second place, relative to a single frame, any physical body set into motion with speed u shortens in the direction of its length, as was postulated by Fitzgerald and Lorentz, in the ratio $\sqrt{1 - u^2/c^2}$. In one sense, the contraction is perhaps not a "real" one, since in a frame in which it is at rest the body measures the same as before; but, as far as effects on surrounding bodies are concerned, the contraction is as real as if it were due to a drop in temperature. For example, a row of similar crystals placed in contact, and then accelerated equally so as to preserve the spacing of their centers, would separate because of the contraction under discussion and would allow light to pass through between them. Perhaps we might say that we have here a sort of *kinematical perspective*, analogous in a way to the ordinary observation that an object appears to change in size as it recedes into the distance.

Unfortunately, these effects are too small to observe on the laboratory scale. It can be said, however, that, according to the theory of relativity, the negative result of the Michelson-Morley experiment, for an observer relative to whose frame of reference the apparatus is in motion, is due to the contraction in question.

There is a somewhat similar effect, also, in *time*. Consider any good "clock," by which is meant a physical system containing a periodic motion or process which can be used to mark off equal intervals of time; to make it a "good" one, let the period of this process be always the same when measured in a frame in which the system is momentarily at rest. According to the first postulate of relativity, a vibrating atom will constitute such a clock. Let the clock be at rest in S' , and let an interval equal to its period in this frame be

$$t'_2 - t'_1 = T_0.$$

Then x' is constant at the clock; hence the corresponding interval and period in S will be, from one of Eqs. (60),

$$T = t_2 - t_1 = \gamma(t'_2 - t'_1) = \gamma T_0.$$

This result has again a double meaning. Its most interesting meaning is that the effects of one physical system on another system, relative to which the first system is moving at velocity u , are modified just as if all natural processes on the moving system were slowed down in the ratio $\sqrt{1 - u^2/c^2}$. Thus, let ν_0 be the frequency of some natural process on the moving system, as measured in a frame of reference that is moving with this system and in which the system is, therefore, at rest. Then the frequency of this process, as measured in

a frame relative to which the system is moving with velocity u , is

$$v = v_0 \sqrt{1 - \frac{u^2}{c^2}}. \quad (62)$$

It follows, for example, that spectral lines, when observed with a spectroscope relative to which the atoms emitting the lines are in motion, should show a slight displacement toward the red, as compared with the lines from stationary atoms, this displacement being superposed upon the ordinary Doppler effect. This effect appears actually to have been observed by Ives, working with canal rays in hydrogen.¹

65. The Transformation of Velocities.—The Lorentz transformation leads to important formulas for the transformation of *velocities* from one frame of reference to another. These formulas are easily found by noting that, if a moving entity has a velocity v relative to S when measured by the observer who uses S , or v' relative to S' when measured by the S' observer, with corresponding Cartesian components v_x, v_y, v_z and v'_x, v'_y, v'_z , then

$$v_x = \frac{dx}{dt}, \quad v_y = \frac{dy}{dt}, \quad v_z = \frac{dz}{dt}; \quad v'_x = \frac{dx'}{dt'}, \quad v'_y = \frac{dy'}{dt'}, \quad v'_z = \frac{dz'}{dt'}.$$

On the other hand, from (59)

$$\begin{aligned} dx' &= \gamma(dx - u dt) & dy' &= dy \\ dt' &= \gamma\left(dt - \frac{u}{c^2} dx\right) & dz' &= dz. \end{aligned}$$

From these expressions, one finds by substitution that

$$\begin{aligned} v'_x &= \frac{v_x - u}{1 - uv_x/c^2}, & v'_y &= \frac{v_y}{\gamma(1 - uv_x/c^2)}, \\ v'_z &= \frac{v_z}{\gamma(1 - uv_x/c^2)}. \end{aligned} \quad (63a,b,c)$$

As an example illustrating the mode of deduction, we have

$$v'_x = \frac{dx'}{dt'} = \frac{\gamma(dx - u dt)}{\gamma(dt - u dx/c^2)} = \frac{dx/dt - u}{1 - u dx/c^2 dt} = \frac{v_x - u}{1 - uv_x/c^2}.$$

By means of these equations we can finally verify that the Lorentz transformation as we have written it makes the speed of light in vacuum always equal to c . For from (63a,b,c) we find, for the square of the velocity v' ,

$$v'^2 = v_x'^2 + v_y'^2 + v_z'^2 = \frac{1}{(1 - uv_x/c^2)^2} \left[(v_x - u)^2 + \left(1 - \frac{u^2}{c^2}\right)(v_y^2 + v_z^2) \right],$$

¹ Ives and Stilwell, *J. O. S. A.*, vol. 28, 215 (1938).

since $1/\gamma^2 = 1 - u^2/c^2$. But $v_x^2 + v_y^2 + v_z^2 = v^2$, or $v_y^2 + v_z^2 = v^2 - v_x^2$. Hence, expanding $(v_x - u)^2$ in the bracket, we obtain

$$\begin{aligned} v'^2 &= \frac{1}{(1 - uv_x/c^2)^2} \left[v^2 - 2uv_x + u^2 - \frac{u^2}{c^2} (v^2 - v_x^2) \right] \\ &= \frac{1}{(1 - uv_x/c^2)^2} \left[v^2 - 2uv_x + \frac{u^2 v_x^2}{c^2} + u^2 \left(1 - \frac{v^2}{c^2} \right) \right]. \end{aligned}$$

If now $v = c$, the expression in brackets reduces to

$$c^2 \left(1 - \frac{2uv_x}{c^2} + \frac{u^2 v_x^2}{c^4} \right) = c^2 \left(1 - \frac{uv_x}{c^2} \right)^2,$$

so that $v'^2 = c^2$ and $v' = c$. In a similar way it can be shown that $v = c$ if $v' = c$.

If, on the other hand, $v < c$, then so is v' , and vice versa, as can readily be shown.

The equation given above for the transformation of velocities measured in *different frames* should not be confused with the ordinary rules for the composition of two velocities measured in the *same frame*. The latter rules are still valid, of course. To take a numerical example, let two electrons, ejected from a filament stationary in S , move off with equal speeds of magnitude $0.9c$, one going toward $-x$ and the other toward $+x$. Then their speed relative to each other, still measured in S , is $1.8c$, by the usual rule. This exceeds c . But, if we make $u = -0.9c$, so that the frame S' keeps up with the electron going toward $-x$, by (63a) the velocity of the second electron relative to the first, *measured now in S'* , is

$$\frac{[0.9c - (-0.9c)]}{(1 + 0.9^2 c^2/c^2)} = \frac{1.8c}{1.81},$$

which is a little less than c .

If in Eqs. (63a,b,c) we let $c \rightarrow \infty$, so that $\gamma \rightarrow 1$, the equations approach the form that is familiar in Newtonian kinematics:

$$v'_x = v_x - u, \quad v'_y = v_y, \quad v'_z = v_z.$$

At the same time the equations of the Lorentz transformation, Eqs. (59), pass over into the Newtonian ones as given in Sec. 58. Thus the Newtonian relations constitute an approximation that is valid as long as we are dealing with velocities much below that of light.

66. General Significance of the Theory of Relativity.—Before proceeding further with the technical discussion of the theory, it may be worth while to pause here for a few final remarks concerning its

general significance. The new ideas involved in it were felt to be revolutionary by the physicists of 30 years ago. They had been familiar with relativistic reversals in the realm of space relations; for example, the left side of a road for one man is the right side for another, and "up" for us is "down" to an Australian (on the opposite side of the earth). But it had seemed that time was time, the same for everybody. Now they were asked to believe that for one observer it is possible that event *A* may precede event *B*, whereas for another *B* precedes *A*, and there is no way of deciding that one observer is right and the other one wrong. It does seem strange, too, that light should move (in a vacuum) with a constant velocity determined only by itself; and yet, if we measure its velocity *relative* to some frame of reference, we obtain always the same value, no matter how fast the frame may be moving in an attempt to keep up with the light. One is tempted to wonder how nature manages to accomplish these remarkable feats.

Probably we should recognize that we are simply being misled here by our ordinary experience, which does not include tossing light beams back and forth. Perhaps, also, the student of the present is not troubled by these things as were the older physicists a generation ago. It may even be that to the modern student the truth of Einstein's relativity postulate will almost be self-evident.

A word may be added, too, concerning the status of the "ether." This word has disappeared almost completely from the language of research physicists; they speak only of electromagnetic "fields" or the like. Certainly an ether that cannot be said to move relative to anything else leads a tenuous existence. The deathblow to the ether, however, has perhaps been struck by the discovery of the photon properties of light. For a shower of photons obviously needs no ether to provide for its propagation.

67. Relativistic Mechanics. The Variation of Mass.—The laws of mechanics as they left the hands of Newton are found upon examination to be not quite in harmony with the new theory of relativity. The necessary corrections to them were discovered originally in studying the motion of charged particles in electromagnetic fields, and this still constitutes the most illuminating path of approach. To follow the argument, however, requires a considerable knowledge of electromagnetic theory; and, on the other hand, the proper relativistic corrections can also be inferred without difficulty from a study of ordinary mechanical phenomena.

For this purpose, we select for study a phenomenon the outcome of which can be inferred from considerations of symmetry, so that

it cannot be in doubt. Let two exactly similar elastic balls approach each other along parallel lines and at equal speeds relative to a frame of reference S' , and let them collide with each other elastically. Then, by the law of conservation of energy, their total kinetic energy must be the same after the impact as before; and, because of the symmetry, this requires that the balls rebound with their speeds unaltered and with their directions of motion deflected through the same angle (Fig. 34). In describing the collision, let Cartesian axes be drawn so that all lines of motion make equal angles with the x -axis and are parallel to the xy plane.

Now let us describe this same collision in terms of another frame of reference, S , which is moving toward $-x$ relative to S' at speed $-u$.

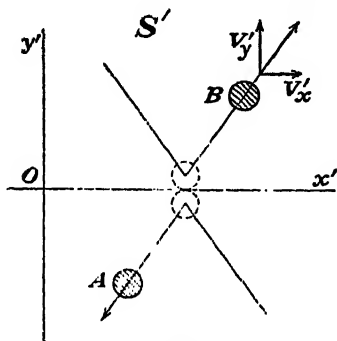


FIG. 34.—Diagram of a symmetrical collision of two similar balls.

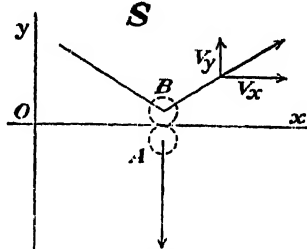


FIG. 35.—The same collision viewed from another frame of reference.

Axes being drawn in the usual manner, let the x' -components of velocity of the two balls relative to S' be $-v'_x$ and $+v'_x$ respectively; and for simplicity let us take $-u = -v'_x$. Then one ball, which we may call A , relative to S , moves up in a direction parallel to the y -axis and rebounds with a simple reversal of its motion, whereas the other, which we may call B , rebounds obliquely (Fig. 35).

In terms of either frame, the x -component of the velocity of both balls remains constant; the y -component, on the other hand, is exactly reversed by the collision. As measured in S' , let A change from a y -component v'_y to $-v'_y$; then, because of the assumed symmetry, B changes from $-v'_y$ to v'_y . According to Newtonian mechanics, the same statement holds in terms of the S frame; but the requirements of relativity make this impossible. In terms of S , let A change in the collision from w_y to $-w_y$ and B from $-v_y$ to v_y . Then, by applying (63b) to the motion first of A and then of B , we obtain the two equations,

$$v_y' = \frac{w_y}{\gamma}, \quad v_y' = \frac{v_y}{\gamma(1 - uv_x/c^2)},$$

v_x being the x -component of the velocity of B referred to S .
Dividing these two equations, we find that

$$\frac{v_y}{w_y} = 1 - \frac{uv_x}{c^2}. \quad (64)$$

Thus $v_y < w_y$, so that, as viewed from S , ball B suffers a smaller change in its y -component of velocity than A .

This result makes it clear that the mechanics of Newton as it stands is not in harmony with the new theory of relativity. It needs some sort of correction. A hint as to the probable form of this correction is furnished by the theory of the electromagnetic field. In Sec. 35 it was found possible to preserve the principle of the conservation of momentum, but the momentum of a *charged* body could not be supposed to be exactly proportional to its velocity, as it is in Newtonian mechanics. The *inertial mass* of a charged body, defined as the ratio of its momentum to its velocity, was found to *increase with the velocity*, because of the electromagnetic momentum in the field. This suggests that we should endeavor here to preserve the principle of the conservation of momentum but should allow the inertial mass to vary.

Let us assume, therefore, that in the collision just described the balls undergo equal and opposite changes in their y -components of *momentum*, defining momentum in the usual way as the product of inertial mass and velocity. Then, if m_A , m_B are the respective inertial masses of the two balls relative to frame S , we have

$$2m_A w_y = 2m_B v_y;$$

hence

$$\frac{m_B}{m_A} = \frac{w_y}{v_y},$$

or, by (64),

$$\frac{m_B}{m_A} = \left(1 - \frac{uv_x}{c^2}\right)^{-1}.$$

Here, as stated in the beginning, $u = v_x'$, v_x' being the x -component of the velocity of B referred to S' . Hence by (63a)

$$u = v_x' = \frac{v_x - u}{1 - uv_x/c^2}.$$

Multiplying this equation through by $v_x (1 - uv_x/c^2)$, we obtain

$$uv_x - \frac{u^2 v_x^2}{c^2} = v_x^2 - uv_x,$$

therefore

$$1 - \frac{v_x^2}{c^2} = 1 - 2 \frac{w_x v_x}{c^2} + \frac{w_x^2 v_x^2}{c^4} = \left(1 - \frac{w_x v_x}{c^2}\right)^2.$$

Hence

$$\frac{m_B}{m_A} = \left(1 - \frac{v_x^2}{c^2}\right)^{-1/2}. \quad (65)$$

To see what this result implies, let $w_y \rightarrow 0$, $v_y \rightarrow 0$. Then m_A approaches a limit m_0 , which is known as the *rest mass* or the mass of the ball for its initial motion out of a state of rest. The other ball B , on the other hand, moves in the limit parallel to x with speed $v = v_x$; and by (65) its mass m_B approaches a limiting value m given by

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (66)$$

In this equation, m_0 represents the mass of B when it is at rest, the two balls being by hypothesis just alike.

Thus the theory of relativity requires that the inertial mass of a body shall vary with its velocity v in proportion to $(1 - v^2/c^2)^{-1/2}$. If this is assumed, it can be shown that the principle of the conservation of momentum can be preserved for all types of elastic collision.¹ The law of variation thus inferred from the theory of relativity is the same as that actually found experimentally to hold for the mass of the electron. Unfortunately the increase in mass is so small at velocities attainable in the laboratory that it has not been detected for any other body. Thus Newtonian mechanics represents an approximation to the correct mechanics which is close enough for all ordinary purposes.

The momentum of a body of mass m moving with speed v can then be written in either of two forms, thus:

$$p = mv = \frac{m_0 v}{(1 - v^2/c^2)^{1/2}}. \quad (67)$$

68. Force and Kinetic Energy.—All other mechanical terms and principles need now to be scrutinized in order to see whether or not

¹ The student can easily verify this statement in the case of the collision studied above, when $v_y \neq 0$. The total velocity of A being $w = w_x$ and that of B , $v = (v_y^2 + v_x^2)^{1/2}$, it is only necessary to show that

$$m_0 w \left(\frac{1 - w^2}{c^2} \right)^{-1/2} = m_0 v \left(\frac{1 - v^2}{c^2} \right)^{-1/2},$$

which verifies the conservation of y -momentum. The x -momentum of both balls remains unchanged.

changes are required to bring them into harmony with the theory of relativity.

The *force* on a body may be defined as the *rate of change of its momentum*. This definition is valid in Newtonian mechanics. If the mass varies, however, and if we retain this definition of force, we are compelled to abandon the familiar $F = ma$; for in linear motion we have then

$$F = \frac{d}{dt}(mv) = m \frac{dv}{dt} + v \frac{dm}{dt} = ma + v \frac{dm}{dt}.$$

Because of this complication, it is usually best to attack problems in relativistic mechanics in terms of momentum rather than of acceleration.

The definition of *work as force times distance* and of *energy as stored work* can be retained without change. This procedure is in harmony with electromagnetic theory. The new expression for the *kinetic energy* T of a moving body can then be found in the usual way by calculating the work done in setting it moving. This gives

$$T = \int_{v=0}^{v} F ds = \int_{v=0}^{v} \frac{d}{dt}(mv) \frac{ds}{dt} dt = \int v \frac{d}{dt}(mv) dt = \int v d(mv),$$

since $ds/dt = v$, the instantaneous velocity. Inserting the value of the mass m from Eq. (66), we have then

$$T = \int v d(mv) = \int v d \frac{m_0 v}{(1 - v^2/c^2)^{1/2}} = m_0 \int v \left[\frac{1}{(1 - v^2/c^2)^{1/2}} + \frac{v^2/c^2}{(1 - v^2/c^2)^{3/2}} \right] dv = m_0 \int_0^v \frac{v dv}{(1 - v^2/c^2)^{3/2}} = m_0 c^2 \left[\frac{1}{(1 - v^2/c^2)^{1/2}} \right]_0^v.$$

Thus we find, for the *kinetic energy* of a body of which the rest mass is m_0 , when moving with speed v ,

$$T = m_0 c^2 \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right). \quad (68)$$

We can also expand in powers of v , obtaining

$$\left(1 - \frac{v^2}{c^2} \right)^{-1/2} = 1 + \frac{1}{2} \frac{v^2}{c^2} + \frac{3}{8} \frac{v^4}{c^4} \dots,$$

so that

$$T = \frac{1}{2} m_0 v^2 + \frac{3}{8} m_0 \frac{v^4}{c^2} \dots$$

Thus, if $v \ll c$, T reduces approximately to the ordinary value for the kinetic energy, $\frac{1}{2} m_0 v^2$, as, of course, it must. Under the same cir-

cumstances, the momentum can also be written, as usual, m_0v . In general, Newtonian mechanics constitutes an approximate form of mechanical theory that is valid for any motion which is slow as compared with the speed of light.

69. A Relation between Mass and Energy.—Combining Eqs. (66) and (68), we can write

$$T = (m - m_0)c^2. \quad (69)$$

Thus the kinetic energy of a moving body equals c^2 times its gain in mass due to the motion. This relation suggests that we may think of the increase in energy as the actual cause of the increase in mass. As we have seen, the momentum present in an electromagnetic field can easily be accounted for in terms of a similar idea. It is then an attractive hypothesis to suppose that even the rest mass m_0 is due to the presence of an internal store of energy of amount m_0c^2 . This may be called the *rest energy* of the body.

The total energy of a moving body would then be $m_0c^2 + T$ or

$$W = mc^2 = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}; \quad (70)$$

and we can write for its inertial mass

$$m = \frac{W}{c^2} \quad (71)$$

and for its momentum

$$p = mv = \frac{Wv}{c^2}. \quad (72)$$

For some purposes, it is useful also to have a relation between W and p that does not involve v . Inserting $v = pc^2/W$ from (72) into (70), rationalizing, and rearranging, we obtain

$$W^2 = m_0^2c^4 + p^2c^2. \quad (73)$$

The foregoing relations suggest that inertial mass may be a property of *energy* rather than of matter as such, each erg of energy possessing, or having associated with it, $1/c^2$ gram of mass. The law of the conservation of mass would then become merely another aspect of the law of the conservation of energy.

These ideas have proved very useful in dealing with nuclear phenomena, which will be described in a subsequent chapter. It appears to be possible, for instance, for a γ -ray photon, impinging upon a nucleus, to be converted into an electron and positron. In such a case, the energy of the photon reappears in part as the kinetic energy of the

particles but largely as their rest-mass energy $2m_0c^2$. Some authors prefer to speak of a conversion of energy into mass in such cases; they would say that energy by itself is not conserved, but only the sum total of mass times c^2 and energy. The view stated here seems to be preferable, but the difference is only a matter of words.

The association of mass and energy is not limited to kinetic energy or to the rest mass. It can be shown that relativity requires mass, or at least momentum, to be associated even with potential energy. This is easily shown in special cases; it may be worth while to give an example.

Suppose two equal masses, moving with equal and opposite velocities along the x -axis as seen by an S observer, collide with each other, a spring acting as a buffer between them; just as they come to rest, let a lock snap shut and hold them thereafter combined into a single mass. In this process, the initial kinetic energy of the masses is converted by the collision into potential energy of the spring. Let us view this collision from a second frame of reference, S' , that is moving with velocity u parallel to x . Then, if the velocities of the two bodies before the collision, as measured in S , are $v_1 = v_{x1} = v$, $v_2 = v_{x2} = -v$, the same velocities as measured in S' are, by Eq. (63a),

$$v'_1 = v'_{x1} = \frac{v - u}{1 - \frac{uv}{c^2}} = -u + \left(1 - \frac{u^2}{c^2}\right) \frac{v}{1 - \frac{uv}{c^2}},$$

$$v'_2 = v'_{x2} = \frac{-v - u}{1 + \frac{uv}{c^2}} = -u - \left(1 - \frac{u^2}{c^2}\right) \frac{v}{1 + \frac{uv}{c^2}}.$$

Therefore,

$$\frac{v'_1}{\left(1 - \frac{v'^2_1}{c^2}\right)^{1/2}} = -\frac{u}{\left(1 - \frac{v'^2_2}{c^2}\right)^{1/2}} + \left(1 - \frac{u^2}{c^2}\right) \frac{v}{\left(1 - \frac{u^2 + v^2}{c^2} + \frac{u^2v^2}{c^4}\right)^{1/2}},$$

the last term being obtained in this form after inserting for v'_1 under the radical the first expression given for v'_1 above. Similarly,

$$\frac{v'_2}{\left(1 - \frac{v'^2_2}{c^2}\right)^{1/2}} = -\frac{u}{\left(1 - \frac{v'^2_2}{c^2}\right)^{1/2}} - \left(1 - \frac{u^2}{c^2}\right) \frac{v}{\left(1 - \frac{u^2 + v^2}{c^2} + \frac{u^2v^2}{c^4}\right)^{1/2}}.$$

The total momentum before collision is, therefore,

$$\frac{m_0v'_1}{\sqrt{1 - v'^2_1/c^2}} + \frac{m_0v'_2}{\sqrt{1 - v'^2_2/c^2}} = -m_0u \left(\frac{1}{\sqrt{1 - v'^2_1/c^2}} + \frac{1}{\sqrt{1 - v'^2_2/c^2}} \right). \quad (74)$$

On the other hand, if we suppose the rest mass of the combined body after collision to be merely the sum of the rest masses of the separate bodies, or $2m_0$, the total momentum *after* collision is

$$\frac{-2m_0u}{\sqrt{1-u^2/c^2}}$$

by Eq. (67), since the velocity of the combined body is then $-u$. This is *not equal* to the momentum *before* collision, as can be seen very easily in case v'_1 and v'_2 are both either greater or less than u .

Thus conservation of momentum fails if only the rest masses are taken into account. But now suppose we include in the mass of the combined body the mass that is to be associated, in the manner described above, with that part of the kinetic energy relative to S' which has been converted by the collision into potential energy of the spring. Then we have a total amount of mass which is proportional to the total energy; and, in consequence of the conservation of energy, this total mass is a constant. The total mass after the collision is, therefore, the same as it was before the collision or

$$\frac{m_0}{\sqrt{1-v_1'^2/c^2}} + \frac{m_0}{\sqrt{1-v_2'^2/c^2}}.$$

If we multiply this value of the total mass by the common velocity of the bodies after the collision, which is $-u$, we obtain for the total momentum *after* collision exactly the same expression as that given in Eq. (74) for the total momentum *before* collision. The principle of the conservation of momentum thus holds here if, and only if, we assume that the potential energy in the spring makes its proper contribution to the mass and momentum of the system.

70. Relativity and Electromagnetism.—Contrary to the situation in mechanics, a review of the laws of the electromagnetic field shows that these particular laws are in harmony with relativity just as they stand. This might have been anticipated in view of the fact that the theory of relativity developed out of experiments in that part of the field of electromagnetism which is called "optics."

The distinction between the *electric* field and the *magnetic* field becomes, however, from the new point of view, in part a *relative* one, depending upon the frame of reference that is being employed. This conclusion, at first sight surprising, can be reached by means of elementary considerations. For example, if there is a set of electric charges at rest relative to the S frame, they will produce, as determined with the use of this frame, only an electrostatic field. But to an observer using an S' frame in motion relative to S , these same charges

will constitute current elements and will be surrounded by a magnetic field as well. There must exist general formulas, therefore, for the transformation of electromagnetic fields from one frame of reference to another. We shall not take space to deduce these formulas here, but we may at least write them down.

As measured by the S observer, let the components of the electric field strength, in electrostatic units, be E_x, E_y, E_z , and let those of the magnetic field strength, in electromagnetic units, be H_x, H_y, H_z ; let the corresponding quantities as measured by the S' observer be indicated by the same symbols with the addition of primes. As usual, let S' have a velocity u relative to S , and let $\gamma = 1/\sqrt{1 - u^2/c^2}$, c

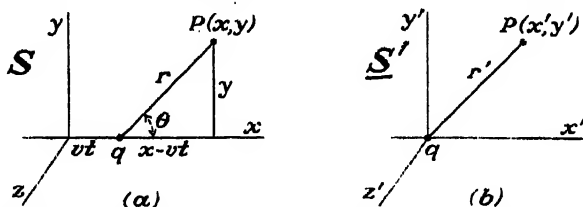


FIG. 36.—Diagram of a point charge moving relative to S , stationary relative to S' .

being the velocity of light. Then the equations of transformation for the electromagnetic vectors are

$$\begin{aligned} E'_x &= E_x, & H'_x &= H_x, & \gamma &= \frac{1}{\sqrt{1 - u^2/c^2}} \\ E'_y &= \gamma \left(E_y - \frac{u}{c} H_z \right), & H'_y &= \gamma \left(H_y + \frac{u}{c} E_z \right), \\ E'_z &= \gamma \left(E_z + \frac{u}{c} H_y \right), & H'_z &= \gamma \left(H_z - \frac{u}{c} E_y \right). \end{aligned}$$

Isolated charges measure the same in all frames, however; this must be so, since the charge on an electron must be a universal constant, and the number of electrons in a body, like the number of distinct objects in any given group, must be the same in all frames of reference. Furthermore, the usual rules for the *force* on a moving or stationary charge still hold; they furnish, in fact, the definition of what is meant by E and H .

The formulas just written may be used to ascertain the effect of uniform motion upon the field of a point charge and thereby to deduce certain formulas that were stated without proof in Chap. II.

To obtain some of these formulas, let a charge q be moving with speed v , and take the x -axis through the charge and in the direction of its motion; draw the xy plane through any point P in the field (Fig. 36). First let us use an S' frame of reference in which the charge is at

rest [Fig. 36(b)]. In this frame the field is purely electrostatic and is already known; if we take the origin at the charge and let x', y', z' be the coordinates of P referred to S' , the components of the field at P are, by Coulomb's law,

$$E'_x = \frac{qx'}{r'^3}, \quad E'_y = \frac{qy'}{r'^3}, \quad E'_z = 0, \quad r'^2 = x'^2 + y'^2. \quad (75a,b,c,d)$$

We now change to an S frame, relative to which S' has a velocity parallel to x of magnitude u . Relative to S , the charge has a velocity $v = u$. Then, from the second set of the above formulas for the transformation of fields, since $H'_x = H'_y = H'_z = 0$, we find, changing u to v ,

$$H_x = 0, \quad H_y = -\frac{v}{c}E_x, \quad H_z = \frac{v}{c}E_y. \quad (76a,b,c)$$

Also, inserting these values of H_y and H_z in the first set of the above formulas, with u replaced by v ,

$$E'_x = E_x, \quad E'_y = \gamma \left(1 - \frac{v^2}{c^2} \right) E_y = \frac{1}{\gamma} E_y, \\ E'_z = \gamma \left(1 - \frac{v^2}{c^2} \right) E_z = \frac{1}{\gamma} E_z. \quad (77a,b,c)$$

Thus, at P , $E_x = 0$, since $E'_x = 0$; and hence, by (76b), $H_y = 0$. Equation (76c), $H_z = (v/c)E_y$, then confirms Eq. (28) in Chap. II.

Also, substituting in (77a,b,c) for E'_x , E'_y from Eqs. (75a,b) and likewise

$$x' = \gamma(x - vt), \quad y' = y$$

from the Lorentz transformation or Eqs. (59), we find

$$E_x = \frac{\gamma q}{r'^3}(x - vt), \quad E_y = \frac{\gamma q y}{r'^3}, \quad r'^2 = \gamma^2(x - vt)^2 + y^2.$$

But vt is the x -coordinate of q , which is at the origin in S' . Hence, if r is the distance of P from q and if θ is the angle between the line qP and the x -axis (cf. Fig. 36a),

$$r^2 = (x - vt)^2 + y^2, \quad x - vt = r \cos \theta, \quad y = r \sin \theta.$$

Thus

$$r'^2 = x'^2 + y'^2 = \gamma^2(x - vt)^2 + y^2 = r^2 + (\gamma^2 - 1)(x - vt)^2 \\ = r^2[1 + (\gamma^2 - 1)\cos^2\theta],$$

and the expressions just found for E_x and E_y can be written

$$E_x = \frac{\gamma \cos \theta}{[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}}} \frac{q}{r^2}, \quad E_y = \frac{\gamma \sin \theta}{[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}}} \frac{q}{r^2}. \quad (78a,b)$$

Thus $E_y/E_x = \tan \theta$, which shows that the electric field points radially outward from the instantaneous location of q . Furthermore,

$$E = (E_x^2 + E_y^2)^{\frac{1}{2}} = \frac{\gamma}{[1 + (\gamma^2 - 1) \cos^2 \theta]^{\frac{3}{2}}} \frac{q}{r^2}, \quad H_z = \frac{v}{c} E \sin \theta$$

from [76c]. These last two equations were cited without proof as Eqs. (31a) and (32) in Chap. II.

In similar fashion, Eq. (34a) in Chap. II, for the radiation field emitted by any charge in linear accelerated motion, can be deduced from Eq. (33) in that chapter, which gives the field for a charge undergoing acceleration out of a state of rest.

Many other conclusions can be drawn by transforming fields from one frame to another; limitation of space allows only the mention of one other. The quantity $E^2 - H^2$ is an "invariant," i.e., for any field, $E'^2 - H'^2 = E^2 - H^2$, as the student can easily verify for himself. This invariance opens the possibility for plane waves in space to transform into plane waves when the frame of reference is changed; for, in such waves, $E = H$, $E^2 - H^2 = 0$. The direction of propagation and the frequency, however, are both altered by the transformation, in accordance with the phenomena of aberration and of the Doppler effect.

71. General Theory of Relativity.—In considering the bearing of the special theory of relativity upon physical laws, we have said nothing about one of the simplest of physical phenomena, *viz.*, *gravitation*. After publishing the special theory, Einstein took up the problem of harmonizing the laws of gravitation with the requirements of that theory. Since, to make the laws of relativity possible, we must suppose that no physical effect can be transmitted from one place to another with a velocity exceeding that of light, it may be assumed that gravitational effects are propagated with a finite velocity. What, then, is the law of this propagation?

At the same time, another idea was active in Einstein's mind. In the special theory, only *unaccelerated* frames of reference had been compared. Why this limitation? Could not the principle of relativity be generalized somehow so that *frames of all sorts* would stand on an equal footing?

In studying these questions, Einstein came to be impressed by the fact that gravitational acceleration is exactly the *same for all bodies*, however much they may differ in density or in other properties. In this respect, gravitational acceleration resembles the *relative* acceleration which appears when a frame of reference is itself subjected to acceleration. The latter effect is a matter of common experience. Every one knows how, when riding on an elevator, he seems momentarily to become lighter whenever the elevator is accelerated downward and heavier when it is accelerated upward. This effect simulates closely an actual change in the force of gravity. By no mechanical experiment, indeed, can an apparent gravitational field thus produced by acceleration of a frame of reference be distinguished from a true field due to gravitational attraction. This fact constitutes a practical difficulty in the blind navigation of airplanes, since it makes impossible the construction of a device to indicate the true vertical unaffected by accelerations of the airplane when in curved flight.

Eventually¹ Einstein came to the conclusion that, in the neighborhood of any given point, there should be *no difference of any kind* between a *gravitational field* due to attracting matter and the "*apparent*" field due to acceleration of a frame of reference. This proposition he adopted as a *postulate* called the *principle of equivalence*. If the principle is accepted, it leads to the prediction of a number of physical effects hitherto unobserved.

Light, for example, had not commonly been supposed to be subject to gravitational action. But suppose the earth's gravitational field were abolished within a laboratory by allowing the whole laboratory to fall freely. Then, relative to the laboratory, there would be no gravitational attraction; a ball thrown out horizontally would travel in a straight line relative to the laboratory, not in a parabola. By the principle of equivalence, therefore, a ray of light projected horizontally would also appear to travel in a straight line; for conditions relative to the laboratory are the same as they would be out in space far from all attracting masses, and there is no doubt that in such locations rays of light are straight. Relative to the earth, however, the path of the ray of light would be slightly curved.

In view of the principle of equivalence it can be said that we ourselves perceive a gravitational field on the earth only because we are using the wrong frame of reference. We ought to use a frame relative to which the earth is accelerated upward at the rate g ; using such a frame, we would find that the apparent gravitational field had com-

¹ Cf. EINSTEIN, A., "Origins of the General Theory of Relativity," Jackson, Wylie and Co., Glasgow, 1933.

pletely disappeared. Gravitational influence thus consists merely in determining what class of frames it is relative to which there is no apparent field, and relative to which free bodies move in straight lines. Such frames are called, in a refined sense of the terms, *inertial* or *Galilean* frames of reference.

It does not follow, however, that the gravitational influence of one piece of matter on another is entirely illusory. For only a *uniform* gravitational field can be transformed away *in its entirety* by a proper choice of the frame of reference. In the neighborhood of a single point, any field can be transformed away; but, in general, the choice of frame that does this varies from point to point. For example, relative to a freely falling frame in New York there is no gravitational field in New York, but there is one of double strength in Australia. The inertial frames in New York and in Australia, respectively, have an acceleration relatively to each other of $2g$.

72. Einstein's Law of Gravitation.—In view of the preceding discussion, there evidently remains for consideration the problem as to the law according to which gravitating matter determines just which frames have the inertial property. The law must be such that its consequences agree with those derived from Newton's law of gravitation as a first approximation, since this law describes the motions of the solar system with high accuracy; and it must also be in harmony with the special theory of relativity. Since in attacking this problem both accelerated and unaccelerated frames must be dealt with, Einstein concluded that the law of gravitation could probably be stated most simply in terms of some formulation that would permit the use, not only of any frame of reference in the ordinary sense, but of any sort of coordinates in four-dimensional space-time. With the aid of the mathematician Grossmann, he found out how to write physical laws in a form that is valid *for any choice of space-time coordinates whatever*. The method involved the use of general tensor analysis, which is too complicated for any example of it to be given here. Suffice it to say that Einstein found that, among all possible guesses as to the correct law of gravitation, one stood out in contrast to all others as the simplest in mathematical form. Therefore he adopted this law as a tentative hypothesis¹ and then proceeded to look for predictions based on it which could be tested by experiment.

From the new law of gravitation thus obtained, Einstein deduced three novel effects that might be accessible to observation:

1. Rays of light passing close to a heavy body should be bent toward it. In the case of the sun, the deflection should be inversely

¹ EINSTEIN, *Ann. d. Physik*, vol. 49, p. 769 (1916).

proportional to the distance of closest approach of the ray to the sun's center and, for a ray just grazing the sun's surface, should amount to 1.75 seconds of arc. Stars seen near the sun, for example, during an eclipse, should appear to be displaced outward by this angular amount.

2. Physical processes in a region of low gravitational potential, when compared with similar processes at a point of high potential, should be found to take place more slowly. Consequently, atomic vibrations on the sun should appear to be slowed down, and spectral lines observed in the spectrum of sunlight should be shifted slightly toward the red as compared with lines emitted or absorbed by the same elements on the earth.

3. The motion of the planets should be very slightly altered. In particular, the perihelion of the orbit of Mercury should be caused to precess about the sun at the rate of 43 seconds a century. The effect on the other planets would scarcely be detectable.

It appears that all three of these effects actually occur. In the case of Mercury, calculation shows that perturbations by the other planets should cause an advance of the perihelion by more than 500 seconds a century, but the observed advance is just 43 seconds greater than can be accounted for in this manner. Thus Einstein's theory removed an annoying discrepancy in astronomical theory.

The new theory of space, time, and gravitation thus arrived at is known as the *general theory of relativity*. According to this theory, the spatial behavior of matter is not quite Euclidean. If a triangle of astronomical size near a heavy body like the sun were surveyed by means of rigid rods, with or without the help of light signals, the angles would not quite add up to 180° ; and so on. We have no space to pursue this fascinating subject further, however, nor to discuss its astronomical applications, which are connected with the question, not yet answered with certainty, as to the assumption that should be made concerning the outlying parts of space. Is space finite or infinite? Is it expanding, as the velocities of the nebulae suggest?

For further study of such questions, the student must be referred to other books.

CHAPTER V

THE ORIGIN OF THE QUANTUM THEORY

The quantum theory, first proposed by Planck in 1900, arose out of the inability of the classical physics to explain the experimentally observed distribution of energy in the spectrum of a black body. We have seen in Chap. III how, likewise, the older theory of radiation could not explain the experimentally observed facts of the photoelectric effect. In the present chapter, we shall first discuss certain observed phenomena of radiation which any theory must explain. We shall then attempt to show (1) just how far the problem of black-body radiation can be solved by classical methods based first upon thermodynamics and then upon classical statistical mechanics, and (2) at precisely what point each of these classical methods failed and the introduction of the concept of quanta seemed to offer the only solution to the problem.

The phenomenon of black-body radiation is, however, of relatively minor importance in experimental physics. If the student does not wish to spend time upon it, at least not at this point, he can without any difficulty at all skip this chapter and proceed at once to the next.

73. Thermal Radiation.—It is a matter of common observation that bodies when heated emit radiant energy—or, more simply, radiation—the *quantity* and *quality* of which depend, for any given body, on the temperature of that body. Thus, the rate at which an incandescent lamp filament emits radiation increases rapidly with increase in the temperature of the filament; and the quality of the radiation, as observed visually, changes markedly as the temperature increases—the emitted light being “whiter” at higher temperatures. Radiation, the quantity and quality of which, emitted by any given body, depends *solely* on the temperature of that body, is called *thermal* radiation. It is a characteristic of thermal radiation that, when dispersed by a prism or other similar means, a *continuous* spectrum is formed. In order that thermal radiation may become visually observable, it is necessary that the temperature of the radiator should be 500 to 550°C. or above. Furthermore, thermal radiation is emitted, ordinarily, only by solids or liquids.

There are many types of radiation other than thermal radiation. Thus, gases and vapors, when suitably “excited,” emit a *characteristic*

radiation, which, when dispersed, results in a line or discontinuous spectrum, the wave length of the several lines being characteristic of the emitting substance. The excitation may result from thermal agitation, electric discharge, bombardment by electrons, or absorption of incident radiation of suitable wave length. As examples may be mentioned the sodium flame, the mercury arc, and the luminescence of sodium vapor when illuminated by light of the *D* lines. Characteristic radiation, in the X-ray region of the spectrum, is emitted also by solids and liquids when bombarded by electrons of suitable speeds. Such bombardment results, in addition, in the emission of a *continuous* X-ray spectrum. Finally, certain solids and liquids emit a characteristic radiation when illuminated (or excited) by light of suitable wave length, even though the intensity of the incident light is not sufficient to produce a perceptible rise of temperature of the emitter. This phenomenon is called *fluorescence* if the emission ceases as soon as the exciting light is removed and *phosphorescence* if the emission persists an appreciable time after removing the excitation.

The classification of radiation into the several types—thermal, characteristic, fluorescent, etc.—is based, of course, upon superficial peculiarities in laboratory methods of production and study rather than upon any *real differences* in the nature of the ultimate emitting mechanisms involved in the several cases. What these differences are, *if any*, we are not now in a position to state. As a matter of fact, considerable progress in the study of the problems of radiation can be made without inquiring either into the atomic mechanism involved or into the nature of radiation itself.

In later chapters we shall discuss some of these other types of radiation, but in the present chapter attention will be restricted to *thermal* radiation. We shall call this simply “radiation,” the adjective being omitted to save repetition. A complete account of thermal radiation will not be attempted, the student being referred for this to the standard treatises on heat; only those features will be discussed which bear on the fundamental problem that led to the quantum theory.

74. Some Fundamental Concepts and Definitions. (a) *Total Emissive Power*.—The rate at which a given body emits radiation depends upon the temperature of the body and on the nature of its *surface*. We may define the *total emissive power*, symbol E , of a body as the “total radiant energy emitted per unit time per unit area of surface of the radiating body.” The total emissive power E increases rapidly with increasing temperature and may be conveniently expressed in ergs per square centimeter per second. Thus, the total emissive

power of cast iron at a temperature of $1600^{\circ}\text{K}.$ ¹ is about 1.1×10^8 ergs per square centimeter per second; that of tungsten at $2450^{\circ}\text{K}.$ is about 5×10^8 ergs $\text{cm.}^{-2} \text{sec.}^{-1}$ (or 50 watts cm.^{-2}).

(b) *Monochromatic Emissive Power.*—If the radiation from a heated body, such as an incandescent lamp filament at a given temperature, be dispersed into a spectrum by a prism or other suitable device, it will be found, by means of a sensitive thermopile, that the energy in the spectrum is distributed among the various wave lengths in a regular manner, as is shown by the curve in Fig. 37, which shows the distribution of energy in the spectrum of tungsten at a temperature of $2450^{\circ}\text{K}.$ We may define the *monochromatic emissive power* e_{λ} at any

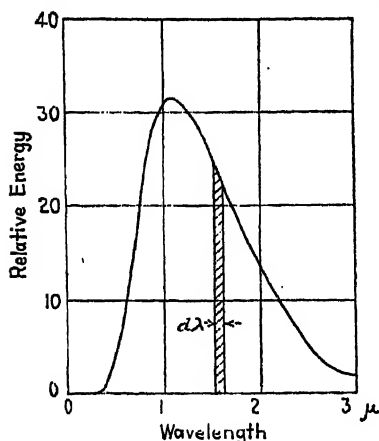


FIG. 37.—Distribution of energy in the spectrum of a tungsten lamp.

given wave length λ by saying that the radiant energy emitted in the spectral range λ to $\lambda + d\lambda$, per unit area per unit time, is given by $e_{\lambda} d\lambda$. Thus, the ordinates of the curve in Fig. 37 are e_{λ} , and the area of the shaded strip is $e_{\lambda} d\lambda$. Since the total emissive power E refers to *all* wave

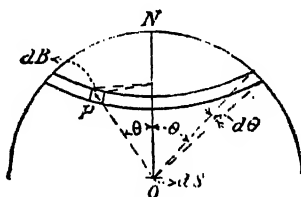


FIG. 38.

lengths combined, we obviously have the relation between E and e_{λ} given by

$$E = \int_0^{\infty} e_{\lambda} d\lambda.$$

That is, the total emissive power is proportional to the area between the curve and the wave-length axis.

The terms "total emissive power" and "monochromatic emissive power" are not to be confused, respectively, with "total emissivity" and "monochromatic emissivity," which we shall define later (Sec. 76).

(c) *Intensity of Radiation from a Surface.*—It will be convenient, to avoid confusion, to define the term *intensity of radiation* from a surface. Let dS (Fig. 38) be a small surface element of a radiating body. Describe about dS a hemisphere of radius ρ and let dB , located

¹ "K." refers to the absolute Kelvin scale of temperature, in centigrade degrees.

at P , be a small element of the surface of this hemisphere, the radius OP to this element making an angle θ with the radius ON which is normal to dS . The rate dQ/dt at which radiant energy is incident upon dB is easily seen to be proportional to (1) the area dB , (2) the area dS of the radiator, and (3) to $1/\rho^2$, because of the inverse-square law. It will depend also upon θ . The apparent area of dS as seen from P , or the area of the projection of dS upon a plane perpendicular to OP , is $dS \cos \theta$. It might be thought, therefore, that dQ/dt would also be proportional to $\cos \theta$; and this well-known "cosine law" is, in fact, approximately obeyed by most surfaces. We can write, therefore,

$$\frac{dQ}{dt} = i dS \cos \theta \frac{dB}{\rho^2}, \quad (79)$$

where i is a quantity that is nearly or quite constant for a given radiator at a given temperature. But dB/ρ^2 is the element of solid angle $d\omega$ which the area dB subtends at O . Hence i is defined by the equation

$$i = \frac{dQ/dt}{dS d\omega \cos \theta}.$$

The coefficient i is called the (total) *intensity of radiation* from dS in the direction OP ; it is the rate at which dS radiates energy in a given direction per unit solid angle and per unit of its own area as projected on a plane perpendicular to the given direction.¹

The definition just stated applies to the *total* radiation from the surface. As in the case of emissive power, we may designate by "monochromatic intensity of radiation," symbol i_λ , the intensity of radiation in a wave-length range $d\lambda$ at wave length λ .

When i is the same in all directions, there is a very simple relation between it and the total emissive power E . Referring to Fig. 38, we may choose as an element of area of the hemisphere a ring of width $\rho d\theta$ and of length $2\pi\rho \sin \theta$, the area of the ring being, thus, $2\pi\rho \sin \theta \rho d\theta$. The flux of energy dQ/dt through this ring is given, as in Eq. (79), by

$$\begin{aligned} \frac{dQ}{dt} &= i dS \cos \theta \frac{2\pi\rho(\sin \theta)\rho d\theta}{\rho^2} \\ &= 2\pi i dS \cos \theta \sin \theta d\theta. \end{aligned}$$

If this equation is integrated from $\theta = 0$ to $\theta = \pi/2$, we shall have the total rate of flux of energy dQ/dt from dS . We have designated by E

¹ The projection of a plane surface dS on a plane making an angle θ with dS has an area $dS \cos \theta$, since all lines on dS that are perpendicular to the intersection of the two planes are shortened by projection in the ratio $\cos \theta$, whereas lines parallel to the line of intersection are not shortened at all.

the total rate of flux of energy per unit area from a surface. Therefore,

$$E = \frac{dQ/dt}{dS} = 2\pi i \int_0^{\pi/2} \cos \theta \sin \theta d\theta = \pi i \sin^2 \theta \Big|_0^{\pi/2}.$$

Therefore

$$E = \pi i. \quad (80)$$

And similarly we have

$$e_\lambda = \pi i_\lambda.$$

(d) *Absorptivity*.—In general, radiation falling upon a surface is partly absorbed, partly reflected, and, unless the body be very thick or very opaque, partly transmitted. We shall define the *absorptivity* of a surface, symbol A , as the fraction of the radiant energy, incident on the surface, which is absorbed. Absorptivity is (1) a pure numeric, (2) for any actual body, less than unity, and (3) varies greatly with wave length of the incident radiation and, to a lesser extent, with the temperature of the absorber.

As we shall see in Sec. 76, there is a very simple relation between the total emissive power E of a surface and its absorptivity A .

(e) *Reflectivity*.—We may define the reflectivity of a surface, symbol R , as the fraction of the radiant energy incident upon the surface, which is reflected. R is a pure numeric.

(f) *Transmissivity*.—Likewise, we may define the transmissivity, symbol T , of a body as the fraction of the radiant energy, incident on the surface of the body, which is transmitted. We shall, in this chapter, confine our attention to cases where the body is so thick or so opaque as to transmit no energy, *i.e.*, to bodies for which $T = 0$. In these cases, since all the incident radiation is, then, either reflected or absorbed, we may write

$$A + R = 1.$$

(g) *The Density of Radiation*.—The radiant energy per unit volume in a stream of radiation is spoken of as the “energy density of the radiation,” symbol ψ . Thus, the solar constant¹ is about 1.94 calories per square centimeter per minute or 1.3×10^6 ergs per square centimeter per second. This is the amount of energy contained in a column² of solar radiation 1 cm.² in cross section and 3×10^{10} cm. long, *i.e.*, in 3×10^{10} cm.³ The energy density of the sun’s radiation

¹ The solar constant is the amount of the sun’s radiation received on unit area in unit time, the receiving area being perpendicular to the sun’s rays and at a distance from the sun equal to the mean radius of the earth’s orbit.

² The column being chosen in the neighborhood of the earth’s orbit, and the sun’s rays being considered parallel throughout the length of the column.

in the neighborhood of the earth is, therefore,

$$\psi = \frac{1.3 \times 10^6}{3 \times 10^{10}} = 4.3 \times 10^{-5} \text{ erg per cubic centimeter.}$$

In the case of the sun's rays, just considered, the radiation is streaming in parallel directions. A different condition exists in the interior of a hollow, heated enclosure. Here radiation is streaming back and forth in all possible directions. The term "energy density" then refers to the total quantity of this radiation in unit volume of the interior.

If a surface obeys the cosine law, so that the intensity i of the radiation emitted by it is the same in all directions, then there is a simple relation between the energy density in the emitted radiation and the total emissive power E of the surface. For, referring again to Fig. 38, let the radius ρ of the sphere be made very large. Then all rays from points on dS to points on dB may be regarded as parallel. The radiant energy emitted in 1 second from dS , in the direction of the radius to dB , will be contained, therefore, in a cylinder whose length is the velocity of light c (vacuum being assumed). The cross-sectional area of this cylinder is $dS \cos \theta$ (this being the area of the projection of dS on a plane perpendicular to the radius); hence its volume is

$$dV = c dS \cos \theta.$$

The total amount of this radiant energy is dQ/dt times 1 second or, by Eq. (79),

$$dE = i dS \cos \theta \frac{dB}{\rho^2}.$$

Dividing dE by dV , we have for the energy density due to this radiation

$$d\psi = \frac{i dB}{c \rho^2}.$$

Since i is assumed to be uniform and since $\int dB = 2\pi\rho^2$ (or half the area of the sphere), the total radiation density is thus

$$\psi = \int d\psi = \frac{i}{c \rho^2} \int dB = \frac{2\pi i}{c}.$$

Comparing this equation with (80) above, we see that the total radiation density is

$$\psi = \frac{2E}{c}. \quad (81)$$

This conclusion applies to the total radiation. But we may also define as "monochromatic energy density," symbol ψ_λ , a quantity such that $\psi_\lambda d\lambda$ is the amount of energy present in unit volume in the form of radiation in a wave-length range $d\lambda$ at wave length λ . Then clearly

$$\psi_\lambda = \frac{2e_\lambda}{c} \quad (82)$$

75. The Black Body and the Isothermal Enclosure.—The maximum possible value of the absorptivity for any surface is unity, which would mean that the surface absorbs all of the radiation falling upon it. No body having this property occurs in nature. Some bodies, however, such as black velvet or lampblack, reflect only a very small fraction of the incident radiation. Accordingly, an ideal body which absorbs all of the radiant energy incident upon it is called the ideal *black body*.

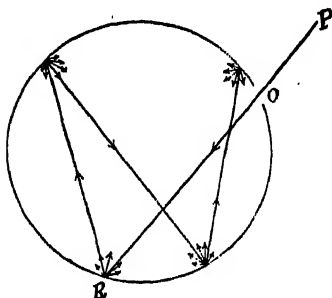


Fig. 39.—The absorption of energy by an artificial black body.

We can approximate to a black body as closely as we please by making use of a small hole in the side of a hollow enclosure. Thus, in Fig. 39, let a ray of radiant energy PR enter the small hole O in the side of a hollow sphere, say of iron. A large part of the radiation is absorbed as it strikes the inside of the sphere at R , the remainder being

diffusely reflected. Only a very small fraction of this diffusely reflected energy goes out through the hole, the rest being finally completely absorbed by successive reflections. Theoretically, perfect absorption would be reached only when the area of the hole is infinitely small compared with the area of the hollow interior. Practically, an approximation sufficient for experimental purposes is obtained by using a hole 1 or 2 cm. in diameter in the end of a hollow cylindrical tube some 20 cm. long and 4 or 5 cm. in diameter.

When such an enclosure is heated, the inside walls radiate, and some of this radiation passes out through the hole. The question arises, whether or not this *emitted* radiation is the same as that which would be emitted by the surface of a black body at the temperature of the enclosure. To answer this question, we may note first that the radiation issuing from the hole will be practically the same as the radiation falling upon an equal area of the walls of the enclosure, provided the hole is very small. Thus we are led to study the distribution of radia-

tion inside of a closed cavity whose walls are all at the same temperature. Such a cavity is called an *isothermal enclosure*.

It can be shown theoretically that the *radiation inside an isothermal enclosure* must possess some very simple properties. In such an enclosure, *the stream of radiation in any given direction must be the same as in any other direction; it must be the same at every point inside the enclosure; and it must be the same in all enclosures at a given temperature. Furthermore, all of these statements hold for each spectral component of the radiation taken separately.*

The proof of these statements proceeds by showing that, if any one of them were not true, it would be possible to construct a device that would violate the second law of thermodynamics. For example, if the stream of radiation traveling west were greater than that traveling north, we could introduce two similar absorbers, one facing east and the other south. One of these absorbers would then become hotter than the other by absorbing radiant energy from the stronger stream. We could, therefore, operate a Carnot engine, using the two absorbers as source and sink, respectively, and so could convert heat continuously into work without leaving other changes in the system, in violation of the second law. Or, again, if the radiation were stronger in one region of the enclosed space than in another, we could locate one absorber in the first region and the other in the second; or, if two enclosures differed as to the radiation in them, we could put one of the two absorbers in each enclosure. Radiation of a particular wave length could be tested by using selective absorbers. The thoughtful student can easily pick flaws in the proofs as sketched here, but they can actually be made quite rigorous.

A simple relation can now be established between the stream of radiation in an isothermal enclosure and the radiation emitted by a black body. Suppose such an enclosure contains a black body at the same temperature. Then the radiation leaving the surface of this body will consist entirely of radiation emitted by it, since a black body reflects none of the radiation that falls upon it. Hence *the stream of radiation emitted by a black body in any direction is the same as the stream of radiation traveling in one direction in an isothermal enclosure at the same temperature*. Hereafter, we shall refer to both types indiscriminately as black-body radiation.

The *energy density* in the enclosure, however, will be twice that due to emission from a black body. For the radiation emitted by the body is confined to a hemisphere, whereas in the enclosure radiation is traveling in all directions. Hence, for the total energy density ψ , or the monochromatic energy density ψ_λ , in the enclosure, we can

write, from Eq. (81) or (82),

$$\psi = 4\frac{E}{c}, \quad \psi_\lambda = 4\frac{e_\lambda}{c}, \quad (83a,b)$$

E being the total emissive power or e_λ the monochromatic emissive power of a black body at the same temperature.

The question left open above can now be answered. The radiation that issues from a small hole in the wall of an isothermal enclosure is exactly the same as that which would be emitted by a black surface of the same size as the hole and at the temperature of the enclosure. This fact is made use of in practice. To study black-body radiation, an electric furnace is employed, consisting of a long tube, preferably (but not necessarily) with blackened walls, heated by an electric current flowing in a wire wound around the tube. The temperature of the central part of the interior is measured with some sort of thermometer. A small hole is made through the wall, and the radiation issuing through it is observed. This radiation approximates very closely that which would be emitted by a black body at the temperature of the furnace.

Black-body radiation is a phenomenon of great interest from the theoretical standpoint, because its properties have a universal character, being independent of the properties of any particular material substance. Two questions concerning it press at once for an answer. How does the energy density in black-body radiation vary with the temperature? And what is the spectral distribution of the radiation? Furthermore, we wish to understand how this particular distribution is brought into existence by the atomic processes going on in matter.

Concerning the first two questions, it was found possible, during the last century, to obtain further information from thermodynamics *without making any assumption as to the atomic process*. The method consisted in considering the effect of expanding or contracting an isothermal enclosure and taking account of the work done on the walls by the radiation in consequence of *radiation pressure*. The radiation was thus treated as a sort of radiation gas. In order to follow the argument, we must ascertain the relation between the pressure and the energy density in uniformly distributed radiation. Before doing this, however, we may first note in passing one other useful conclusion that can be drawn from the properties of the isothermal enclosure.

76. Relation between Emissive Power and Absorptivity.—In an isothermal enclosure, every element of surface must be emitting as much radiation as it absorbs, otherwise its temperature would change. Suppose two surfaces S_1 and S_2 have emissive powers E_1 and E_2 and

absorptivities A_1 and A_2 , respectively. Then, since the same stream of radiation falls on both, S_2 will be absorbing A_2/A_1 as much radiation as S_1 . It follows that S_2 must also be emitting A_2/A_1 as much radiation as S_1 . Hence

$$\frac{E_2}{E_1} = \frac{A_2}{A_1} \quad (84)$$

Thus, for different surfaces at the same temperature, *emissive power and absorptivity must be proportional*. This conclusion can be shown to hold for each wave length separately.

If, in particular, we make $A_1 = 1$, so that S_1 is a black surface, E_1 has obviously the maximum value that is possible at the given temperature; for A_1 in the last equation cannot exceed unity. Thus no surface can emit more strongly than a black body. If E_0 is the emissivity of a black body and E the emissivity of any other body whose absorptivity is A , we find, upon putting in the last equation $E_1 = E_0$, $A_1 = 1$, $E_2 = E$, $A_2 = A$,

$$E = AE_0. \quad (85)$$

These conclusions have all been confirmed by experiment, as set forth in treatises on heat. The relations expressed by Eqs. (84) and (85) are commonly known as "Kirchhoff's law."

Note.—There is considerable confusion in the terminology of the emission and absorption of radiation. Thus, for the quantity which we have herein defined as "total emissive power," one finds used, variously, such terms as "radiating power," "emission," and "emissivity." And for the quantity which we have called "absorptivity," one finds "absorbing power" and "absorptive power." The use of the word "power" in "total emissive power" is consistent with its use in mechanics. For "power" in the technical sense, in physics and engineering, means a "rate of doing work" or of delivering energy; and a radiating surface is emitting energy at a definite rate. But "power" used in connection with absorption, as in the term "absorbing power," is clearly inappropriate; for the quantity *defined* does not involve a *rate* of absorbing energy.

One must distinguish between "absorptivity," as herein defined, and "coefficient of absorption," which is an entirely different term, referring to the absorption of radiation in its passage *through* matter. Thus, if a beam of radiation of intensity I_0 is incident upon and passes through a slab of absorbing material, the thickness of which is d , the intensity I of the emerging beam is given by the equation $I = I_0 e^{-\mu d}$, where "e" is the Napierian base of logarithms and μ is the *coefficient of absorption* of the material.

It is essential, also, to distinguish between the terms "total emissive power" and "emissivity." If E is the total emissive power of a body at temperature T , and E_0 is the total emissive power of a black body at the same temperature, we have seen that $E = AE_0$, where A is the absorptivity of the substance. Now the emissivity of a surface is defined as "the ratio of its total emissive power to the total emissive power of a black body at the same temperature"; i.e., emissivity equals

E/E_0 . Thus the emissivity of a substance is a pure numeric and is equal to its absorptivity. The same relations hold between monochromatic emissive power and the corresponding absorptivity.¹

77. Pressure Due to Isotropic Radiation.—Suppose a stream of radiation in a vacuum falls normally on the surface of a body. Then, if w is the mean energy density in the oncoming waves, they carry also w/c units of momentum per unit volume [Eq. (25b) in Sec. 35]. Thus the waves bring up to each unit area of the surface, along with cw ergs of energy [Eq. (21b) in Sec. 34], w units of momentum per second, the momentum as a vector being directed normally toward the surface. If the waves are absorbed by the surface, it receives this momentum and experiences, therefore, a pressure equal to w . Similarly, if a surface emits a stream of radiation in the normal direction, with average energy density w , this stream carries away momentum directed away from the surface equal to w units per second, and there must be a reaction on the surface equivalent to a pressure of magnitude equal to w .²

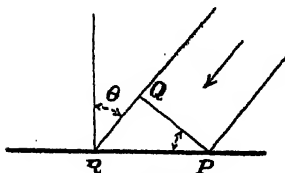


FIG. 40.

Suppose, however, that the radiation, instead of falling normally on the surface, is incident at an angle θ . Then the energy that crosses a unit area drawn perpendicular to the rays (PQ in Fig. 40) is received by an area of magnitude $1/\cos \theta$ on the surface (PR in Fig. 40). Furthermore, the component of the momentum normal to the surface is less than in the case of normal incidence in proportion to $\cos \theta$. Thus the momentum in the direction of the normal that is delivered to unit area of the surface per second is decreased by the obliqueness of incidence in the ratio $\cos^2 \theta$, and the resulting pressure, if the radiation is entirely absorbed, is

$$p = w \cos^2 \theta, \quad (86)$$

w being, as before, the energy density in the incident radiation. The same expression holds for the pressure caused by the emission of a beam at an angle θ , or for the additional pressure caused by the occurrence of a reflected beam. If an incident beam in which the energy density is w is specularly reflected from a surface at the angle of incidence θ , the total pressure on the surface is $2w \cos^2 \theta$.

¹ So far as possible, the terminology defined in the "International Critical Tables," vol. I pp. 34–42, has been adopted in this chapter.

² Pressure and energy density are easily seen to be physical quantities of the same dimensions.

Finally, let radiation be streaming toward a surface and also away from it with equal intensities in all directions, as in an isothermal enclosure. Such a distribution of radiation is equivalent to a large number of beams of plane waves, all of equal intensity, with their directions of propagation distributed equally in direction. Let there be N beams in all, and let the energy density due to any one of them be w . Then the total energy density just in front of the surface and the pressure p on it are

$$\psi = Nw, \quad p = \Sigma w \cos^2 \theta = w \Sigma \cos^2 \theta, \quad (87a, b)$$

by (86), $\Sigma \cos^2 \theta$ denoting the sum of the values of $\cos^2 \theta$ for all of the beams.

To find this latter sum, imagine lines drawn outward from a point O on the surface to represent the various directions of the beams, whether moving toward the surface or away from it, and then about O as center draw a hemisphere of unit radius with its base on the surface

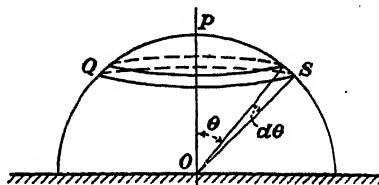


FIG. 41.

(cf. Fig. 41, where only two of the lines are shown). From the surface of the hemisphere cut out a ring-shaped element of area, QS , by means of two cones of semiangle θ and $\theta + d\theta$, drawn from O as apex and with the normal OP as axis. The edge of this element is a circle of perimeter $2\pi \sin \theta$, and its width is $d\theta$, hence its area is $2\pi \sin \theta d\theta$; whereas the area of the whole hemisphere is 2π . Now the lines of approach of the N beams of radiation, if drawn through O , will cut the hemisphere in points equally distributed over its surface. Hence, if we let dN denote the number of these lines that pass through the ring-shaped element, dN will be to N in the ratio of the area of the ring to the area of the hemisphere, whence¹

$$\frac{dN}{N} = \frac{2\pi \sin \theta d\theta}{2\pi} = \sin \theta d\theta.$$

The value of $\cos^2 \theta$ is the same for all of the dN beams. Hence their contribution to $\Sigma \cos^2 \theta$ is, from the last equation,

$$\cos^2 \theta dN = N \cos^2 \theta \sin \theta d\theta.$$

Thus

$$\Sigma \cos^2 \theta = \int \cos^2 \theta dN = N \int_0^{\pi/2} \cos^2 \theta \sin \theta d\theta = \frac{1}{3} N. \quad (88)$$

¹ In more succinct terms, $2\pi \sin \theta d\theta$ is an element of solid angle about O , having the form of a conical shell, and 2π is half of the whole solid angle about O .

(The limit is $\pi/2$ because directions all around the normal OP are included in the ring.) For the pressure we thus obtain, from (87b), $p = \frac{1}{3}wN$, or, by (87a),

$$p = \frac{1}{3}\psi. \quad (89)$$

Thus the average pressure on the walls of an isothermal enclosure is one-third of the energy density in it.

78. Deduction of the Stefan-Boltzmann Law.—In 1884, Boltzmann deduced a theoretical law for the variation of the total intensity of black-body radiation with temperature. For this purpose, he applied the laws of the Carnot cycle to an engine in which the radiation played the part of the working substance.

The ideal Carnot engine consists of a cylinder with walls impervious to heat, a piston likewise impervious to heat and moving without

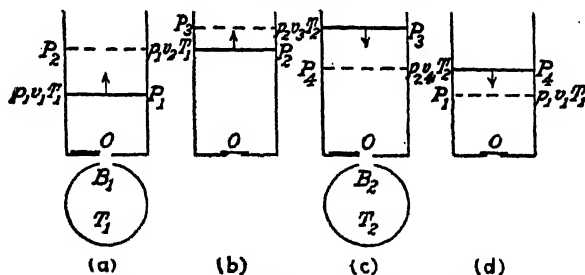


FIG. 42.—Boltzmann's radiation engine.

friction, and a base (end opposite to piston) through which heat may enter or leave. For our purpose, we shall make the walls, piston, and base perfectly reflecting¹ except for a small opening O in the base which may be covered at will by a perfectly reflecting cover. Inside the cylinder is a vacuum. Let this cylinder be placed with the opening O uncovered and opposite an opening in an isothermal enclosure B_1 , which is maintained at temperature T_1 (Fig. 42). Then the cylinder will fill up with radiation entering it through O from B_1 until there is the same density ψ_1 of radiation in the cylinder as there is in B_1 , at which time radiation will be passing at the same rate from O to B_1 as from B_1 to O .

¹ "Perfectly reflecting" surfaces cannot quite be realized in the laboratory, just as walls "impervious to heat" cannot, although good approximations are possible in both cases. It seems probable, however, that the processes responsible for absorption of radiation or for heat conduction are independent of those general properties of matter which give rise to the existence of the general laws of thermodynamics. In any case, the use of idealized surfaces or substances of the kind in question has never led to conclusions in conflict with experiment.

Remembering that when the energy density within the cylinder is ψ , the pressure on the cylinder walls is $\frac{1}{3}\psi$ [see Eq. (89)], we may now consider the following cycle of events:

- a. Starting with the piston in the initial position P_1 [Fig. 42(a)], the initial volume of the cylinder being v_1 and the initial pressure $p_1 = \frac{1}{3}\psi_1$, we cause the piston to move upward, *slowly*, until position P_2 is reached, the volume increasing to v_2 . During this process the radiation density within the cylinder will remain constant at ψ_1 . To keep it constant, additional radiation must enter the opening O from the enclosure B_1 , for two reasons:

1. External work W_e is done by the radiation on the piston. If T_1 remains constant, ψ_1 and, therefore, p_1 likewise remain constant, and

$$W_e = \frac{1}{3}\psi_1(v_2 - v_1).$$

2. The interior of the cylinder has *increased* in volume by $(v_2 - v_1)$, which requires an additional influx of energy equal to $\psi_1(v_2 - v_1)$.

Thus the total influx H_1 of radiation from B_1 must be

$$H_1 = \frac{4}{3}\psi_1(v_2 - v_1). \quad (90)$$

This isothermal process is represented on the p - v diagram (Fig. 43) by the line P_1P_2 . The energy H_1 is equivalent to heat supplied to the space within the cylinder, just as in an ordinary Carnot cycle the first isothermal expansion is accompanied by an absorption of heat. An amount of heat equal to H_1 must be supplied from external sources to B_1 in order to keep the temperature of B_1 constant.

- b. When the piston has reached P_2 , the perfectly reflecting cover is placed over the opening O [Fig. 42(b)], thereby effecting complete thermal isolation of the interior of the cylinder, and a further expansion to position P_3 is made. External work is done, as before, on the piston, the energy required for this external work being supplied by the radiation. Partly because of this work and partly because of the increase in volume, *the energy density of the radiation within the cylinder must decrease* from ψ_1 to some smaller value, ψ_2 . The pressure, likewise, has decreased. This is obviously an adiabatic process. It is represented in Fig. 43 by the line P_2P_3 .

The new energy density ψ_2 is equal to the energy density in an enclosure at a certain new temperature T_2 . If we make the

expansion during this second process very small, we may represent the change in temperature ($T_1 - T_2$) by dT , and the corresponding change in energy density, $\psi_1 - \psi_2$, by $d\psi$. Since $p = \frac{1}{3}\psi$, we have then

$$dp = \frac{1}{3}d\psi, \quad (91)$$

dp representing the change in radiation pressure.

- c. The engine is now placed opposite a second isothermal enclosure B_2 [Fig. 42(c)], at temperature T_2 , the slide is removed from the opening O , and the piston is moved, by the application of suitable external force, from P_3 to P_4 . On account of this compression, there is a tendency for the density of radiation within the cylinder to rise and for radiation to pass through O into B_2 . The compression is supposed to take place so slowly, however, that the radiation density remains constant at a value only infinitesimally in excess of ψ_2 . During this second isothermal process, radiant energy in amount H_2 leaves the engine.
- d. Having reached a suitable point P_4 , the opening O is closed, and the radiation is further compressed adiabatically until the initial position P_1 is reached, corresponding to the initial volume and energy density.

The net external work done during this cycle is represented by the area $P_1P_2P_3P_4$ of Fig. 43. If we assume the change of pressure to have been very small, this area equals $(v_2 - v_1) dp$.^{*} Calling the net external work dW , we have, therefore,

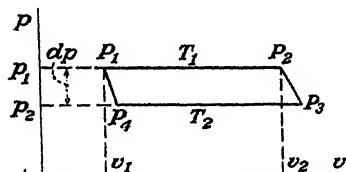


FIG. 43.—The p - v diagram for the Carnot cycle of the radiation engine.

$$dW = (v_2 - v_1) dp = \frac{1}{3}(v_2 - v_1) d\psi$$

by (91). Hence, by the usual rule for a Carnot cycle,

$$\frac{dW}{H_1} = \frac{T_1 - T_2}{T_1} = \frac{dT}{T_1},$$

or, using the value just found for dW and Eq. (90),

$$\frac{\frac{1}{3}(v_2 - v_1) d\psi}{\frac{4}{3}(v_2 - v_1)\psi_1} = \frac{dT}{T_1}.$$

^{*} This is the area of a rectangle of height dp and length P_1P_2 . The small triangular areas at the ends will not be equal, but, as dp approaches zero, these areas become negligible in comparison with the area $P_1P_2P_3P_4$. Hence the difference between the latter area and the rectangle can be ignored in the limit as $dp \rightarrow 0$.

Thus, dropping the subscript, we have quite generally

$$\frac{d\psi}{\psi} = 4 \frac{dT}{T}.$$

This equation integrates to give $\log \psi = 4 \log T + \text{const.}$, or

$$\psi = aT^4, \quad (92)$$

where a is some constant.

Equation (92) states that *the energy density of the radiation within an isothermal enclosure is proportional to the fourth power of the absolute temperature T* . Since the emissive power E of a black body is proportional to the energy density within an enclosure at the same temperature, i.e., $E \propto \psi$, it follows also that *the total emissive power of a black body is proportional to the fourth power of its absolute temperature*. We may write, therefore,

$$E = \sigma T^4 \quad (93)$$

where σ is another constant. This equation is known as the "Stefan-Boltzmann law," the constant σ being known as "Stefan's constant" or as the "Stefan-Boltzmann constant." According to Eq. (83a) above, a and σ are connected by the relation

$$a = \frac{4\sigma}{c}. \quad (93a)$$

79. Experimental Verification of the Stefan-Boltzmann Law.—The importance of the Stefan-Boltzmann law of total radiation has inspired many investigators to check its validity. Among these may be mentioned Lunmer and Pringsheim,¹ who studied the total radiation from an experimental black body over a temperature range from 373 to 1535°abs. Their data are shown graphically in Fig. 44, in which the abscissas are absolute temperatures of the black body as measured with a thermoclement, which, in turn, was calibrated (indirectly) against a nitrogen gas thermometer. The ordinates are the fourth roots of the total emissive power E (in relative units) of the black body. Except for the two lowest points (which involved troublesome corrections), the relation between $\sqrt[4]{E}$ and T is seen to be in excellent agreement with the Stefan-Boltzmann law.

From these measurements, Lunmer and Pringsheim obtained a value for σ of 5.32×10^{-5} . Numerous recent investigations have shown that the true value is considerably higher, the present accepted

¹ *Ann. d. Physik*, vol. 63, p. 395 (1897). These experiments are described in Preston's "Theory of Heat."

value¹ being $\sigma = 5.735 \times 10^{-5}$ erg cm.⁻² sec.⁻¹ deg.⁻⁴. From σ and Eq. (93a) we find $a = 7.652 \times 10^{-15}$ erg cm.⁻³ deg.⁻⁴. From these two constants, one may compute the total emissive power E of a black

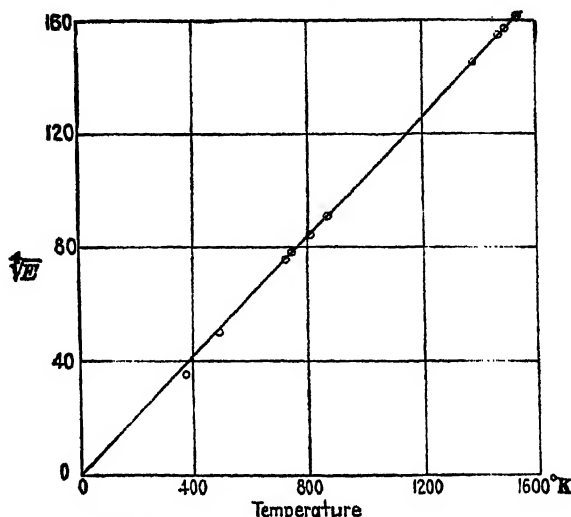


FIG. 44.—Verification of the Stefan-Boltzmann total radiation law.

body at any temperature and the energy density ψ of the radiation within an enclosure by the formulas

$$E = 5.735 \times 10^{-5} T^4 \text{ erg cm.}^{-2} \text{ sec.}^{-1},$$

$$\psi = 7.652 \times 10^{-15} T^4 \text{ erg cm.}^{-3}$$

Thus, the total emissive power of a black body at a temperature of 1000°abs. is

$$E_{1,000} = 5.73 \times 10^7 \text{ ergs cm.}^{-2} \text{ sec.}^{-1}$$

The energy density of the radiation within an enclosure at the same temperature is

$$\psi = 7.65 \times 10^{-3} \text{ erg cm.}^{-3};$$

and the pressure due to radiation on the walls of such an enclosure is 2.55×10^{-3} dyne cm.⁻².

80. Reflection from a Moving Mirror.—In the preceding discussion no attention was paid to the spectral distribution of the radiation. No use was made, in fact, of the wave theory of light. The radiation pressure might easily have the value that was assigned to it above if the corpuscular theory were true. The question presents itself,

¹ BRIDGMAN, *Rev. Modern Physics*, vol. 1, p. 61 (1929).

however, whether or not the Stefan-Boltzmann law can be applied not only to the total radiation but also to the separate wave lengths. In order to investigate this question, we need to know what happens to the spectral distribution of a beam of radiation when it is *reflected from a mirror that is in motion*, such as the piston in the ideal apparatus that was employed in the deduction of the last section.

Let us consider first the effect of such motion upon a monochromatic beam. In Fig. 45(a), two successive wave fronts are shown at an instant when they have been partly reflected from a mirror MM ; the mirror is supposed to be moving with a component of velocity u perpendicular to its plane, u being taken positive when the mirror is receding from the light. Parts of two incident waves are AB and

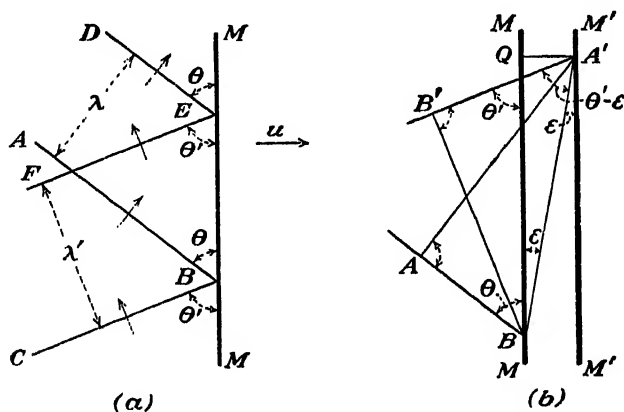


FIG. 45.—Reflection from a moving mirror.

DE , which are 1 wave length or a distance λ apart and are falling on the mirror at an angle of incidence θ ; CB and FE are parts of the same waves that have been reflected and are now leaving the mirror at an angle of reflection θ' , a distance λ' apart. It is obvious from the figure that $\lambda = BE \sin \theta$, $\lambda' = BE \sin \theta'$, whence

$$\frac{\lambda'}{\lambda} = \frac{\sin \theta'}{\sin \theta}. \quad (94)$$

If the mirror is stationary, we have $\theta = \theta'$, hence $\lambda = \lambda'$. To find what the law of reflection is when the mirror is moving, consider *two successive positions of the same wave*. In Fig. 45(b), the part AB of a wave is just beginning to fall on the mirror at an angle of incidence θ , the instantaneous position of the mirror being at that instant MM . The same portion of the wave at a later instant, after it has been reflected at the angle θ' , is shown by $A'B'$, the point A being now in

contact at A' with the mirror, which is in the new position $M'M'$. While A traveled along the ray AA' , B traversed BB' , hence

$$AA' = BB'.$$

The angles BAA' and $BB'A'$, being angles between ray and wave, are right angles. It follows, therefore, by similar triangles, that angle

$$B'A'B = ABA'$$

or

$$\theta' - \epsilon = \theta + \epsilon, \quad (95)$$

where ϵ is the angle QBA' . Furthermore, while A went from A to A' with the speed c of light, the mirror moved from MM to $M'M'$ with the speed u . Hence, if $A'Q$ is a perpendicular dropped from A' onto MM ,

$$\frac{A'Q}{AA'} = \frac{u}{c}$$

or

$$\frac{BA' \sin \epsilon}{BA' \sin (\theta + \epsilon)} = \frac{\sin \epsilon}{\sin (\theta + \epsilon)} = \frac{u}{c}$$

Substituting here $\epsilon = \frac{1}{2}(\theta' - \theta)$, from (95), we obtain

$$\frac{\sin \frac{1}{2}(\theta' - \theta)}{\sin \frac{1}{2}(\theta' + \theta)} = \frac{u}{c}. \quad (96)$$

From this one easily obtains

$$\tan \frac{1}{2} \theta' = \frac{c + u}{c - u} \tan \frac{1}{2} \theta \quad (97)$$

as the law of reflection from a moving mirror. If $u = 0$, it reduces to the familiar law $\theta' = \theta$.

The change of wave length implied by Eq. (94) represents a Doppler effect due to reflection from a moving mirror. It is this change which is needed for the further study of thermal radiation. We shall be dealing only with infinitesimal velocities, however, so that θ' and θ are nearly equal. For this case we can put $\theta = \theta'$ in the denominator in Eq. (96), and we can replace the sine by the angle in the numerator, obtaining

$$\frac{1}{2}(\theta' - \theta) = \frac{u}{c} \sin \theta;$$

and, similarly, subtracting 1 from both sides in (94), we have

$$\frac{\lambda' - \lambda}{\lambda} = \frac{\sin \theta' - \sin \theta}{\sin \theta} = \frac{2 \cos \frac{1}{2}(\theta' + \theta) \sin \frac{1}{2}(\theta' - \theta)}{\sin \theta} = \frac{\theta' - \theta}{\sin \theta} \cos \theta$$

in the limit as $\theta' \rightarrow \theta$. From these two equations we obtain

$$\frac{\lambda' - \lambda}{\lambda} = 2 \frac{u}{c} \cos \theta;$$

whence, writing $\Delta\lambda = \lambda' - \lambda$, we have, for the increase in wave length upon reflection from a mirror receding with indefinitely small velocity u ,

$$\Delta\lambda = 2 \frac{u}{c} \lambda \cos \theta. \quad (98)$$

81. Effect of an Adiabatic Expansion upon Black-body Radiation.—

Returning now to the sequence of operations described in Sec. 78, let us consider the effect of the adiabatic process (*b*) upon the spectral distribution of the radiation. In this process, black-body radiation initially at temperature T_1 imprisoned in a cylinder with perfectly reflecting walls is slowly expanded from an initial energy density ψ_1 to a new energy density ψ_2 . Let the former restriction to a small expansion be dropped. The change in direction of the rays that is produced by the moving piston, according to Eq. (97) above, will tend to make the radiation no longer isotropic. We can obviate this inconvenient effect, however, by letting part of the walls of the cylinder reflect perfectly *but diffusely*. A surface of magnesium oxide does this very well. Then, if the expansion is made very slowly, because all rays (except a negligible few) strike the diffusing surface repeatedly, the radiation will be kept effectively isotropic; and the pressure on the piston, according to Eq. (89) above, will be at all times equal to $\frac{1}{3}\psi$.

Let the cylinder have a cross section A and a (variable) length l . Then, if ψ is the energy density at any moment, when the piston moves outward a distance dl , work $p dV = \frac{1}{3}\psi A dl$ is done on it by the force due to radiation pressure. This work is done at the expense of the enclosed energy, the total amount of which is $lA\psi$. Hence

$$\begin{aligned} \frac{1}{3}\psi A dl &= -d(lA\psi) = -A\psi dl - Al d\psi, \\ \frac{d\psi}{\psi} &= -\frac{4 dl}{l}, \end{aligned}$$

and, after integration,

$$\log \psi = -\log l^4 + \text{const.}, \quad \psi = Cl^{-4}, \quad (99)$$

C denoting a constant.

Now it can be shown by thermodynamical reasoning that an expansion of the type considered here cannot destroy the black-body property of the radiation. For, starting with the cylinder full of black-body radiation at density ψ_1 and temperature T_1 , let us expand

it to a density ψ_2 . Let T_2 be the temperature of an enclosure in which the density has this same value ψ_2 . Suppose that in the cylinder there were more radiation per unit volume of wave lengths near some value λ' than at the same wave lengths in the enclosure, and less radiation near some other wave length λ'' . It would then be possible to cause little radiation to pass from the cylinder into a second enclosure at temperature T'_2 slightly above T_2 , by covering the opening in the base of the cylinder with a plate transmitting wave lengths near λ' but reflecting all others¹ and putting the cylinder into communication with the second enclosure through this opening. In a similar way enough radiation near λ'' could be passed into the cylinder from an enclosure at a slightly lower temperature T''_2 to restore the total energy to ψ_2 . Then the radiation could be compressed back to ψ_1 , the change in ψ and l and the amount of the work done being just the reverse of these quantities during the expansion. Finally, putting the cylinder again into communication with the enclosure at T_1 , we could allow the spectral distribution to be restored to that proper to a black body at

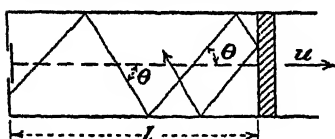


FIG. 46.

T_1 , but without any net transfer of energy between cylinder and enclosure since the total energy density has already been restored to that corresponding to T_1 . Thus we should have performed a cyclic operation the only effect of which

is to transfer heat energy from an enclosure at T''_2 to one at a higher temperature T'_2 . But this is inconsistent with the second law of thermodynamics.

Thus black radiation must remain black during any slow adiabatic expansion. Its density, and therefore its temperature, however, decrease. On the other hand, we can determine from the Doppler effect how the spectral distribution of the radiation changes. In this way, we may arrive at a knowledge of the effect of changes of temperature upon the spectral distribution of black-body radiation.

Suppose, first, that the walls of the cylinder are perfectly reflecting. Then any ray preserves its angle of inclination to the axis of the cylinder in spite of successive reflections (*cf.* Fig. 46) and has, therefore, a constant component of velocity $c \cos \theta$ perpendicular to the piston. Thus the ray strikes the piston $c \cos \theta / 2l$ times a second, l being the length of the enclosed space; and each time, according to (98), the wave length is increased $2u\lambda \cos \theta / c$, u being the speed of the piston. The wave length thus increases at an average rate of

¹ Selective reflection and transmission are well-known optical phenomena.

$$\frac{d\lambda}{dt} = \frac{c \cos \theta}{2l} \frac{2u\lambda \cos \theta}{c} = \frac{u\lambda}{l} \cos^2 \theta \text{ cm. per sec.}$$

This expression for $d\lambda/dt$ varies with θ . If, however, as assumed above, *diffuse* reflection occurs over part of the walls, the waves take turns moving in different directions, and hence they all experience the same *average* change in length. To simplify the calculation of this average change, imagine that each wave moves so as to be perpendicular to one of a very large number N of lines equally spaced in direction. Then, since the radiation remains isotropic, during a second of time any wave must spend $1/N$ second moving in each of the N directions. While moving in one of these directions, the wave making an angle θ with the axis, its wave length increases, according to the last equation, by the amount

$$t \frac{d\lambda}{dt} = \frac{1}{N} \frac{d\lambda}{dt} = \frac{u\lambda}{lN} \cos^2 \theta.$$

The total change in λ in a second is thus

$$\sum \frac{u\lambda}{lN} \cos^2 \theta = \frac{u\lambda}{lN} \sum \cos^2 \theta,$$

Σ indicating the sum for all N directions. But $\Sigma \cos^2 \theta$ was evaluated above and found to equal $N/3$ [Eq. (88) in Sec. 77]. Hence, all waves of length λ increase during the expansion at the average rate

$$\frac{d\lambda}{dt} = \frac{u\lambda}{3l}.$$

Since $u = dl/dt$, this equation leads to

$$\frac{d\lambda}{\lambda} = \frac{1}{3} \frac{dl}{l}, \quad \log \lambda = \log l^{1/3} + \text{const.},$$

$$\lambda \propto l^{1/3}.$$

But, by (99), $l \propto \psi^{-3/4}$; and, also, by the Stefan-Boltzmann law, $\psi \propto T^4$. Thus $\lambda \propto \psi^{-1/4} \propto T^{-1}$. Hence the wave length of any given spectral component of black-body radiation changes during a slow adiabatic expansion in such a way that

$$\lambda \propto \frac{1}{T}. \quad (100)$$

82. The Wien Displacement Law.—The result just obtained leads to an important conclusion concerning *the manner in which the spectral distribution of black-body radiation depends upon the temperature*. Let $\psi_\lambda d\lambda$ denote the part of the energy in unit volume due to waves of

length between λ and $\lambda + d\lambda$, the total energy density being thus $\int_0^\infty \psi_\lambda d\lambda$. Let us fix our attention on a particular spectral range $d\lambda_1$ at λ_1 containing energy $\psi_{\lambda_1} d\lambda_1$ in the black-body radiation at temperature T_1 . Then an adiabatic expansion which lowers the temperature to some new value T_2 changes the wave lengths of this part of the radiation in the ratio T_1/T_2 , λ_1 becoming λ_2 and $\lambda_1 + d\lambda_1$, $\lambda_2 + d\lambda_2$, where

$$\frac{\lambda_2}{\lambda_1} = \frac{T_1}{T_2}, \quad \frac{\lambda_2 + d\lambda_2}{\lambda_1 + d\lambda_1} = \frac{T_1}{T_2}, \quad \therefore \frac{d\lambda_2}{\lambda_2} = \frac{d\lambda_1}{\lambda_1}. \quad (101a, b, c)$$

(Cf. Fig. 47.) The energy originally in $d\lambda_1$ is also decreased, in the same ratio as is the total energy; for we might have started with only the radiation in $d\lambda_1$ present in the cylinder, and then the argument leading up to Eq. (99) would have held for this part of the radiation alone. Hence, ψ_{λ_2} being the new spectral density of energy,

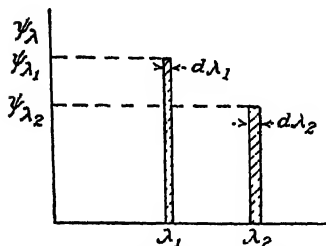


FIG. 47.—Adiabatic change of λ and ψ_λ .

$$\frac{\psi_{\lambda_2} d\lambda_2}{\psi_{\lambda_1} d\lambda_1} = \frac{\psi_2}{\psi_1} = \frac{T_2^4}{T_1^4},$$

by the Stefan-Boltzmann law. Substituting here the value of $d\lambda_2/d\lambda_1$ from (101c), we obtain

$$\frac{\psi_{\lambda_2}}{\psi_{\lambda_1}} = \frac{T_2^5}{T_1^5}. \quad (102a)$$

Since ψ_λ is proportional to the monochromatic emissive power of a black body at the same temperature, we can also conclude that

$$\frac{e_{\lambda_2}}{e_{\lambda_1}} = \frac{T_2^5}{T_1^5}. \quad (102b)$$

Wave lengths λ_2 , λ_1 , such that $\lambda_2/\lambda_1 = T_1/T_2$ may be called *corresponding wave lengths* for black radiation at these temperatures. Equations (102a) and (102b) then state that the *energy density ψ_λ in an isothermal enclosure and the monochromatic emissive power of a black body when taken at corresponding wave lengths are both directly proportional to the fifth power of the absolute temperature*. This important conclusion is known as the *Wien displacement law*.

Equations (102a) and (102b) can be tested experimentally by plotting observed values of ψ_λ/T^5 or of e_λ/T^5 against the product λT . The curve thus obtained should be the same at all temperatures, for at

corresponding wave lengths λT has the same value and so have the ratios ψ_λ/T^5 and e_λ/T^5 .

In Fig. 48 are shown the energy distribution curves obtained for the spectrum of a black body by Lummer and Pringsheim,¹ and in Fig. 49 is shown a composite curve plotted from these data in the manner just described, data taken at three different temperatures being combined as indicated on the plot. The theoretical prediction is seen to be fulfilled.

It is clear that a single curve serves to represent black-body radiation at all temperatures. Or, if we wish to have curves for e_λ or ψ_λ at various temperatures, we can construct them from the known curve at a single temperature. Thus, in Fig. 50, let the solid curve represent experimentally determined values for e_λ at temperature T_1 . Then the curve for any other temperature T_2 can be plotted as follows: Choose any wave length λ_1 , for which the monochromatic emissive power, at temperature T_1 , is e_1 , corresponding to point P_1 on the curve. The wave length λ_2 , which, at temperature T_2 , corresponds to λ_1 , is given by

$$\lambda_2 = \lambda_1 \frac{T_1}{T_2};$$

and the emissive power e_2 , which corresponds to this new wave length at temperature T_2 , is given, by Eq. (102b), as

$$e_2 = e_1 \left(\frac{T_2}{T_1} \right)^5.$$

These two "displacements" shift point P_1 to P_2 , which is a point on the energy distribution curve for temperature T_2 . Thus, to obtain the new curve, abscissas are to be changed in the ratio T_1/T_2 , ordinates in the ratio T_2^5/T_1^5 .

Obviously, if λ_m is the wave length at which e_λ and ψ_λ have their maximum value,

$$\lambda_m T = A = \text{const.}$$

¹ D. phys. Ges. Verhandlungen, vol. 1, pp. 23 and 215 (1899); vol. 2, p. 163 (1900).

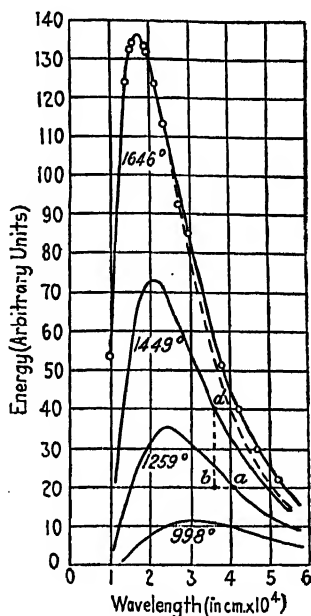


FIG. 48.—Distribution of energy in the spectrum of a black body at various temperatures.

for all temperatures—a special case of the displacement law that is often cited. The present accepted value of the constant A , which is known as Wien's displacement constant, is $A = 0.2884$ cm. deg.

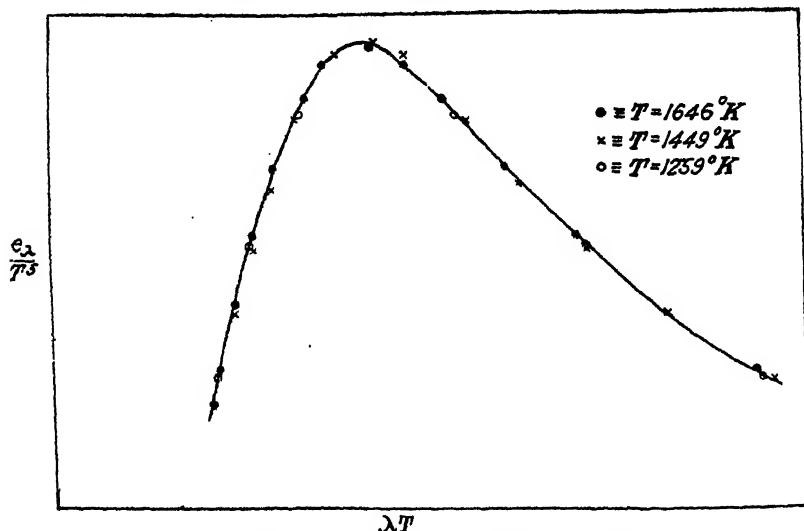


FIG. 49.—Experimental verification of the black-body displacement law.

83. The Formula for Black-body Radiation.—The relationship represented by Fig. 49 can also be expressed mathematically. It is obvious that e_λ/T^5 is some function of the product λT ; we may write

$$\frac{e_\lambda}{T^5} = f(\lambda T).$$

Or, since then

$$e_\lambda = T^5 f(\lambda T) = (\lambda T)^5 \frac{f(\lambda T)}{\lambda^5},$$

we can write

$$e_\lambda = \frac{1}{\lambda^5} F(\lambda T), \quad (103)$$

where $F(\lambda T) = (\lambda T)^5 f(\lambda T)$.

FIG. 50.—Change of the black-body curve with temperature.

Thus, by reasoning based on thermodynamics, the problem of black-body radiation is reduced to the determination of the single unknown function $F(\lambda T)$. In order to determine this function, we must resort to other methods. All attempts to find the function $F(\lambda T)$ on the basis of classical theory, however, failed. Equation (103) represents the limit of success of classical theory in dealing with the problem of black-body radiation.

Two of the formulas that were proposed on the classical basis, although incorrect, deserve some consideration. Wien proposed a formula that was derived from special assumptions concerning the process of emission and absorption. Rayleigh made a suggestion based on very general reasoning,¹ and on the basis of this suggestion Jeans arrived at formulas for ψ and e that were free from all unknown constants. The Rayleigh-Jeans formula was found to fit the experimental curve at very long wave lengths; whereas, by assigning proper values to certain arbitrary constants in Wien's formula, the latter could be made to fit at wave lengths above or near the point of maximum emission. Otherwise, both formulas failed (cf. Fig. 56 in Sec. 93). The method of approach employed by Rayleigh and Jeans is full of interest, and it also furnishes an excellent background for the consideration of Planck's introduction of quanta, by means of which he succeeded in finding the (apparently) *correct* black-body formula. In the next sections we shall consider the reasoning of Rayleigh and Jeans. It is based upon the classical law of the equipartition of energy; therefore this will be discussed first.

84. Degrees of Freedom.—A gas molecule within an enclosure is *free to move* in a variety of ways. It may, in general, have a motion of translation, a motion of rotation, and, *if polyatomic*, the atoms may vibrate with respect to each other. A billiard ball upon the table does not have so great freedom of motion as the gas molecule, for, although the ball has as much freedom of rotation as the molecule, its motion of translation is confined to a plane. A flat disk sliding upon the ice is not so free as is the billiard ball, for the disk is capable of rotation about only one axis, *viz.*, the vertical axis. A block sliding in a closely fitting groove is still more restricted in its freedom of motion.

Now, it is obvious that the relative freedom of motion of these several bodies can be expressed by giving the number of independent quantities which need to be known to express completely the position and orientation of the bodies. Thus, in the case of the block sliding in the groove, its position and orientation are completely specified by one quantity—its distance from some origin. For the disk on the ice, three quantities are required—two to express the position of some point, say the center of gravity of the disk with respect to some coordinate system on the ice, and one more to specify the orientation of the disk with respect to some fixed line. For the billiard ball, five such independent quantities are required. *These separate and independent quantities which need to be known to specify completely the*

¹For an outline of the arguments of Wien and of Rayleigh see Preston's "Theory of Heat."

position and configuration of the body are called its "degrees of freedom."

This concept of degrees of freedom is readily extended to a *system* of bodies. If we have *three* billiard balls on the table, *fifteen* independent quantities are necessary to express the "position and configuration of the system." Degrees of freedom are additive—the total number of degrees of freedom of a system of bodies is the sum of the number of degrees of freedom possessed by the several bodies which make up the system.

To each degree of freedom of a system there corresponds an independent term in the expression representing its kinetic energy as a function of its coordinates. The number of these terms is equal to the number of the degrees of freedom of the system.

For example, a monatomic gas molecule behaves as if it had a motion of translation only. Accordingly, its kinetic energy is completely expressed with respect to an x -, y -, z -coordinate system by the terms

$$\frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2,$$

where v_x , v_y , and v_z refer to the component velocities in the three coordinate directions and m is the mass of the molecule. If there are N monatomic molecules in a given mass of gas, $3N$ such independent terms are necessary to express the kinetic energy of the gas at any given instant. On the contrary, a rigid diatomic gas molecule would have 5 degrees of freedom, 3 for translation (of the center of gravity) and 2 for rotation of the "dumbbell-shaped" molecule about each of a pair of axes at right angles to each other and to the line joining the atoms. If the molecule is not rigid, but the atoms can vibrate with respect to each other, then there is an additional degree of freedom corresponding to this vibration, and the total number is 6. A triatomic molecule will have at least 6 degrees of freedom (unless all atoms lie on a line), 3 for translation and 3 for rotation about 3 mutually perpendicular axes; if vibration of the atoms is also possible, there will be 9 degrees of freedom in all.

85. Relation between Energy per Degree of Freedom and the Temperature.—A simple means for calculating the energy of a system, according to classical theory, is furnished by the principle of the *equipartition of energy*. We shall not attempt to deduce this principle from the more fundamental principles of statistical mechanics, but we shall state it and illustrate its application in several important cases.

Consider a monatomic gas in an enclosure at a uniform temperature T . According to the kinetic theory of gases, the molecules are in

motion in all possible directions with a wide range of velocities, their average kinetic energies being a function of the temperature. The degrees of freedom of such a system may be divided into three groups, corresponding to the three coordinate axes, *viz.*, an x group, a y group, and a z group. The classical theorem of the *equipartition of energy* now says that *each degree of freedom in one group has, on the average, exactly the same amount of kinetic energy as has, on the average, a degree of freedom in any other group*. For example, let us take a census, or "snapshot," of 1,000 of the gas molecules. Each molecule has a component of velocity in each of the coordinate directions x , y , z . Let v_x , v_y , and v_z be these components, and $\frac{1}{2}mv_x^2$, $\frac{1}{2}mv_y^2$, $\frac{1}{2}mv_z^2$ the corresponding kinetic energies. The *average* kinetic energy per molecule \bar{E}_x in the x direction for the 1,000 molecules is obtained by adding the terms $\frac{1}{2}mv_x^2$ for all the molecules and dividing by 1,000, and similarly for the y - and the z -components. The theorem of the equipartition of energy states that

$$\bar{E}_x = \bar{E}_y = \bar{E}_z.$$

Let us designate by \bar{E}_1 the common value of \bar{E}_x , \bar{E}_y , and \bar{E}_z . Then, if in the gas under consideration there are N molecules each with 3 degrees of freedom, the total kinetic energy of the system, due to the random motion of its molecules, is $3N\bar{E}_1$. The same principle holds for more complicated molecules. The mean kinetic energy of a rigid diatomic molecule, for example, according to classical theory, would be $5\bar{E}_1$; of a rigid triatomic one, $6\bar{E}_1$.

The case of "vibratory" degrees of freedom involves a new feature. A vibratory motion becomes possible only if we assume that the 2 atoms which make up the molecule are bound together by forces which, normally, hold the 2 atoms in certain fixed equilibrium positions with respect to some fixed point within the molecule. Vibrations take place when this equilibrium is disturbed. Now, during any simple harmonic motion, it is easily shown that the average amount of *potential* energy is the same as the average kinetic energy \bar{E}_1 . Hence the total amount of energy associated with the vibratory motion, according to classical theory, is $2\bar{E}_1$. The *total* average energy of a *vibrating* diatomic molecule will then be $5\bar{E}_1 + 2\bar{E}_1 = 7\bar{E}_1$.

Now the quantity \bar{E}_1 is a very simple universal function of the temperature. This may be seen as follows:

According to the kinetic theory of gases, the pressure p which a gas exerts on an enclosure is given by¹

$$p = \frac{1}{3}\rho v^2,$$

¹ See any textbook on general physics.

where ρ is the density of the gas and v_s is the "root mean square" velocity of its molecules. Since $\rho = nm$, where n is the number of molecules per cubic centimeter in the gas and m is the mass of each molecule, we have

$$p = \frac{1}{3} nmv_s^2.$$

Assume that we are dealing with a monatomic gas at temperature T . Then the average kinetic energy \bar{E}_K of each molecule of the gas is given by

$$\bar{E}_K = \frac{1}{2} mv_s^2,$$

and we may write, instead of the last equation,

$$p = \frac{2}{3} n\bar{E}_K.$$

But

$$pV_M = RT,$$

where V_M is the volume of a gram-molecule of the gas at temperature T , and R is a universal constant known as the "gas constant," the numerical value of which is 8.315×10^7 ergs per mole per degree, or 1.986 calories per mole per degree.

Eliminating p between the last two equations, we have

$$\frac{2}{3} n\bar{E}_K V_M = RT.$$

The product of n , the number of molecules per cubic centimeter, by V_M , the volume of the gram-molecule, is equal to N_0 , the number of molecules in a gram-molecule:

$$nV_M = N_0.$$

N_0 , known as *Avogadro's number*, is constant for all substances. Its numerical value, as given in Sec. 45 above, is 6.02×10^{23} molecules per gram-molecule. Eliminating V_M between the last two equations and solving for \bar{E}_K , we have

$$\bar{E}_K = \frac{3}{2} \frac{R}{N_0} T = \frac{1}{2} kT \quad (104)$$

where

$$k = \frac{R}{N_0} \quad (105)$$

The numerical value of the new universal constant k is given by

$$k = \frac{8.315 \times 10^7}{6.02 \times 10^{23}} = 1.381 \times 10^{-16} \text{ erg per molecule per degree.}$$

The constant k is known as "Boltzmann's constant," or as the "molecular gas constant."

Equation (104) gives the average kinetic energy of the monatomic molecule which we are considering. As we have seen, such a molecule has 3 degrees of freedom. The *average kinetic energy per degree of freedom*, therefore, is

$$\bar{E}_1 = \frac{1}{2} kT. \quad (106)$$

The theorem of the equipartition of energy thus leads to a very simple way of computing the classical value of the energy of thermal agitation, kinetic or kinetic plus potential, of a system, provided (1) that the system has a large enough number of degrees of freedom and (2) that temperature equilibrium has been reached.

86. Verification of Equipartition of Energy in the Brownian Motion.

These deductions from the theorem of the equipartition of energy have been verified in a very striking manner, both qualitatively and quantitatively, by observations on the Brownian movements. Over 100 years ago, Robert Brown, an English botanist, noted that minute particles (small enough to be visible only through a high-power microscope) in suspension in liquids undergo continuously random motions, almost as if they were alive. After the development of the kinetic theory of matter, these random motions were explained as due to the continuous, though irregular, bombardment of the particles by the molecules of the liquid. In 1905, Einstein developed a theory of the motion of these particles based upon the concept of the equipartition of energy, according to which their mean kinetic energy of agitation should be exactly the same as that of a molecule of the liquid at the same temperature, which, per degree of freedom, is given by Eq. (106). Einstein's equation is as follows:¹

$$\overline{x^2} = 2kUtT,$$

where x denotes the component in a fixed direction of a displacement of the particle in a time t (several seconds); $\overline{x^2}$ denotes the average of x^2 for a large number of such observed values of x ; k is Boltzmann's constant; U is a constant depending upon the viscosity of the medium in which the particle is suspended and on the size of the particle—actually it is the velocity with which the particle moves through the medium when driven by a force of 1 dyne; and T is the absolute temperature. The value of $\overline{x^2}$ can be obtained from observation by watching the particle for a long time in the field of the microscope and noting its position at fixed intervals of time; U can be computed;

¹ For a derivation of this equation, and a discussion of its experimental verification, see E. H. Kennard, "Kinetic Theory of Gases" p. 281, 1938, or R. A. Millikan, "The Electron."

and it is thus possible to determine k experimentally. Actually, the value of k thus determined is combined with the known value of the gas constant R , and a value of Avogadro's number N_0 is thus determined from the Brownian movements [cf. (105) above]. The value of N_0 so obtained agrees, almost within the limits of error of measurement, with that obtained by dividing the faraday by the charge on the electron, thus proving that the law of the equipartition of energy, upon which Einstein's equation is based, is valid for this type of phenomenon, and incidentally verifying, beyond doubt, the kinetic theory of matter.

87. Degrees of Freedom in an Enclosure.—It was suggested by Lord Rayleigh¹ and by Jeans² that this law might be applied to the radiation problem by computing the number of modes of free vibrations in the ether in an enclosure and by assuming that with each mode of vibration, or degree of freedom, there is associated, on the average, $\frac{1}{2}kT$ of kinetic energy and a like amount of potential energy (See 85), or a total energy of kT per degree of freedom.

Any vibrating system is capable, in general, of a great many modes of vibration. Thus, a violin string or an organ pipe may vibrate not only in its fundamental mode but also in a great many overtones. We may think of these several overtones as corresponding to the various degrees of freedom of the vibrating system.

An organ pipe is a one-dimensional system, so far as its standing sound waves are concerned, since the vibrations take place longitudinally and in one direction only—*viz.*, parallel to the axis of the pipe. Standing waves may be set up, also, in a "three-dimensional organ pipe," *i.e.*, in a hollow enclosure such as a large concrete-lined room. In this case, too, we may have a great many possible overtones, or modes of vibration. We shall compute the number of such modes of vibration within a given frequency range, per unit volume of the enclosure for sound waves, and shall then apply the results to the analogous case—standing electromagnetic waves within an enclosure—and shall thus determine the number of degrees of freedom per unit volume of the electromagnetic field, for a given frequency range. This number, multiplied by the energy per degree of freedom, *assuming the law of equipartition*, should give us the *energy density* of the radiation of that particular frequency.

It will be instructive to apply the method of computation first to the one-dimensional case of a stretched string. Let a string of great length L (Fig. 51) be stretched between two supports A and B . The

¹ *Phil. Mag.*, vol. 49, p. 539 (1900).

² *Phil. Mag.*, vol. 10, p. 91 (1905).

condition that standing waves may be set up requires (1) that A and B (*i.e.*, the ends of the string) shall be nodes of motion and (2) that there shall be an integral number of equal loops between A and B . Since each loop is equivalent to one-half of a wave length, we may say that standing waves may occur in the string only for those wave lengths λ defined by

$$\frac{L}{\lambda/2} = i = \frac{2L}{\lambda}, \quad (107)$$

where i is some positive integer. The standing waves may be regarded as due to the superposition of two trains of waves of equal amplitude and wave length running in opposite directions along the string.

For a string of given length, i is equal, also, to the number of possible modes of vibration of the string for wave lengths equal to

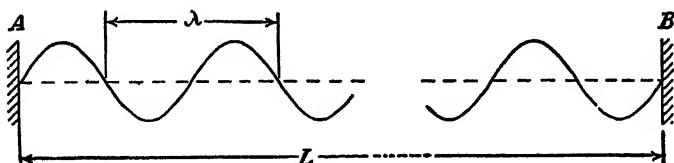


FIG. 51.

or greater than the value of λ as given by (107). We wish to know how many of these waves have lengths within a range $\Delta\lambda$. Now if we increase i by some integer, Δi , the value of λ given by (107) decreases to $\lambda - \Delta\lambda$, where

$$\frac{2L}{\lambda - \Delta\lambda} = i + \Delta i.$$

Subtracting Eq. (107) from this equation, we obtain

$$\Delta i = 2L \left(\frac{1}{\lambda - \Delta\lambda} - \frac{1}{\lambda} \right) = \frac{2L\Delta\lambda}{\lambda(\lambda - \Delta\lambda)}.$$

Since L is very great, however, $\Delta\lambda$, as given by this last equation, must be very small; hence, in the denominator of the last fraction, $\Delta\lambda$ can be neglected in comparison with λ , and we can write

$$\Delta i = 2L \frac{\Delta\lambda}{\lambda^2}.$$

Obviously Δi represents the number of modes of vibration whose wave lengths lie in the range $\Delta\lambda$.

Now, associated with each of these modes of vibration, there are 2 degrees of freedom, since the vibration is transverse and any point

on the string is free to move in a *plane* at right angles to the string. The total number of degrees of freedom per unit length of the string in the wave-length range from λ to $\lambda - \Delta\lambda$ is, therefore,

$$\Delta n = 2 \frac{\Delta i}{L} = 4 \frac{\Delta\lambda}{\lambda^2}. \quad (108)$$

The situation is more complex but involves identically similar principles when we consider the possible systems of standing waves in a hollow enclosure. Let us discuss first the system of sound waves in a rectangular flat box, a sort of two-dimensional organ pipe. Let $ABCD$ (Fig. 52) represent the box, and consider a system of plane-parallel sound waves moving in directions parallel to the top and bottom of the box and being reflected, without absorption, from

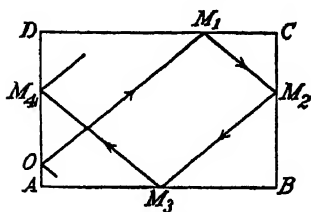


FIG. 52.—Reflection of waves inside a box.

its sides, according to the ordinary law of reflection. A set of waves moving initially in the direction OM_1 will, after reflection at the face CD , move in the direction M_1M_2 ; after reflection at M_2 , the direction M_2M_3 will be parallel but opposite to OM_1 , etc. For this group of waves, only four directions of motion are possible, $\pm OM_1$ and $\pm M_1M_2$.

Under suitable conditions the four trains of waves thus formed will combine to produce a set of *standing waves*. To describe these, divide the sides AB and DC of the box into i_1 equal segments and the sides AD and BC into i_2 segments, as in Fig. 53; and join the points of division by two sets of parallel lines, lines at each point making equal angles with the sides (*e.g.*, $\angle r_3q_2B = \angle s_3q_2C$). Then a group of waves, such as those just described, can be supposed to move so that they are always parallel to one or the other of these two sets of lines. At a given instant, let alternate parallel lines represent, respectively, crests and troughs of these waves. At the instant of observation, suppose a crest moving toward C , for example, coincides with the line p_3r_3 ; as time goes on, this crest becomes extended at the r_3 end by reflection of a crest now at r_3q_2 and in motion toward B and of another crest now at p_3s_1 and moving toward D . A similar crest, s_3q_2 , moving toward C is itself undergoing progressive reflection to form crests p_1s_3 and r_3q_2 ; and so on. Upon these waves, then, let there be superposed a set of waves of equal amplitude moving in the opposite directions, crests moving toward A coinciding momentarily with p_3r_3 and s_3q_2 ; and so on. We obtain in this way four sets of

waves mutually producing each other by reflection at the walls, as in Fig. 52.

The set of waves just described will combine to form standing waves. For after half a period has elapsed, the crests, initially on the solid lines, will be meeting on the dotted lines, whereas troughs will now meet on the solid lines; the result is a disturbance of which the phase is momentarily just opposite to the previous one. After another half period, the original distribution recurs. Surfaces drawn parallel to the sides of the box through the points of subdivision are, like the sides themselves, surfaces of no disturbance. It can be shown that, by proceeding in the manner described, using in turn all pairs of values for i_1 and i_2 , we obtain all possible modes of vibration of the medium inside the box.

For our purpose, the connection between the numbers i_1 and i_2 and the length of the waves is important. In Fig. 53, the wave length is the distance between adjacent parallel solid lines. Let the waves make angles θ_1 and θ_2 with the sides of the box, of which the lengths are l_1 and l_2 , respectively. Then, in Fig. 53, since $r_1 r_2 = r_2 r_3 = l_1 / i_1$,

$$(r_1 r_3) \sin \theta_1 = \frac{2l_1}{i_1} \sin \theta_1 = \lambda,$$

$$(p_1 p_3) \sin \theta_2 = \frac{2l_2}{i_2} \sin \theta_2 = \lambda.$$

Solving these equations for $\sin \theta_1$ and $\sin \theta_2$ in terms of l_1 , l_2 and substituting in the equation $\sin^2 \theta_1 + \sin^2 \theta_2 = \sin^2 \theta_1 + \cos^2 \theta_1 = 1$, we obtain

$$\frac{i_1^2}{l_1^2} + \frac{i_2^2}{l_2^2} = \frac{4}{\lambda^2}.$$

This method of constructing standing waves is readily extended to a *three-dimensional* rectangular box, with edges l_1 , l_2 , and l_3 long. The edges are divided into i_1 , i_2 , and i_3 segments, respectively; and eight sets of waves are drawn. It is then found that instead of the last equation the condition for standing waves is

$$\frac{i_1^2}{l_1^2} + \frac{i_2^2}{l_2^2} + \frac{i_3^2}{l_3^2} = \frac{4}{\lambda^2}. \quad (109)$$

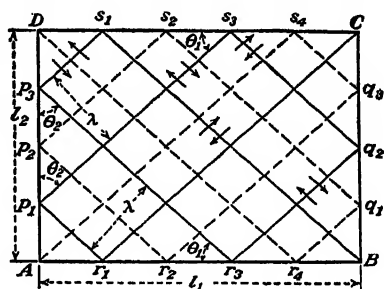


FIG. 53.—Standing waves in a two-dimensional box.

From this equation we can find the number of possible modes of vibration within the box for which the wave length is greater than any given value λ_m . The number of such modes will be equal to the number of possible combinations of positive integers i_1, i_2, i_3 , which make the left-hand member of (109) less than $4/\lambda_m^2$. To find this number, imagine each possible set of the i 's represented on a three-dimensional plot by a point the coordinates of which are $x = i_1/l_1$,

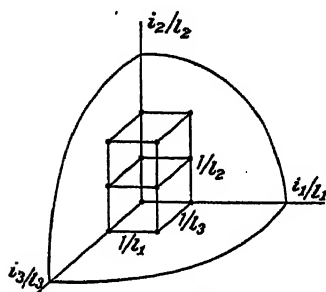


FIG. 54.

$y = i_2/l_2, z = i_3/l_3$ (Fig. 54). These points will lie at the corners of rectangular parallelepipeds or cells the edges of which are $1/l_1, 1/l_2, 1/l_3$ long (two of them being shown in the figure). There are just as many cells as there are points; each cell has eight points at its corners, but, on the other hand, each point serves as a corner to eight cells. Hence, since the volume of a cell is $1/l_1 l_2 l_3$, there are $l_1 l_2 l_3$ cells or points in each unit volume

of the plot. The points we wish to count, according to Eq. (109), lie inside the sphere defined by the equation

$$x^2 + y^2 + z^2 = \frac{4}{\lambda_m^2},$$

of which the radius is $2/\lambda_m$. The number of these points will be equal to the volume of 1 octant of this sphere multiplied by the number of points in unit volume, or to¹

$$\left[\frac{1}{8} \times \frac{4}{3} \pi \left(\frac{2}{\lambda_m} \right)^3 \right] \times l_1 l_2 l_3 = \frac{4\pi}{3} \frac{l_1 l_2 l_3}{\lambda_m^3}.$$

This is the total number of the possible modes of vibration for which $\lambda > \lambda_m$. Dividing it by the volume of the box, which is $l_1 l_2 l_3$, and dropping the subscript m , we have, therefore, for the number of modes of vibration, or of degrees of freedom, per unit volume of the box, corresponding to wave length greater than λ ,

$$n = \frac{4\pi}{3} \frac{1}{\lambda^3}.$$

The number of degrees of freedom per unit volume in a wave-length range λ to $\lambda + d\lambda$ is thus

$$dn = 4\pi \frac{d\lambda}{\lambda^4}, \quad (109a)$$

¹ This is not quite accurate, of course, because of errors at the surface, but, when n is large, the relative error is negligible.

both dn and $d\lambda$ being taken positive for simplicity. This is analogous to Eq. (108) for the string.

In the discussion up to this point we have been dealing with the longitudinal vibrations of sound waves. An identical analysis will hold for electromagnetic waves inside a rectangular enclosure with reflecting walls, with one exception. In the latter case, since these waves are transverse, the number of degrees of freedom is twice as great as for longitudinal waves, corresponding to the two possible independent planes of polarization. For *electromagnetic* waves in an enclosure, therefore, we may write

$$dn = 8\pi \frac{d\lambda}{\lambda^4}, \quad (109b)$$

where dn is the number of degrees of freedom of the system of waves, per unit volume of the enclosure, in the wave-length range λ to $\lambda + d\lambda$.

88. The Rayleigh-Jeans Formula.—We are now in position to apply to the radiation problem the law of the equipartition of energy. Consider the radiation within an enclosure the walls of which are at a temperature T , equilibrium having been reached between the radiation and the walls. Whatever may be the nature of the radiating and absorbing mechanism of the walls, each degree of freedom of the *entire* system, *walls and electromagnetic field included*, should have associated with it an average amount of kinetic energy $\frac{1}{2}kT$, as given by Eq. (106). Each “vibratory” degree of freedom has an equal amount of potential energy. The total amount of energy associated with each degree of freedom in the field should be, therefore, kT . Hence, the number of degrees of freedom per unit volume for the wave-length range λ to $\lambda + d\lambda$ multiplied by kT should give the amount of radiant energy per unit volume in this range, which we denoted previously (Sec. 74) by $\psi_\lambda d\lambda$. Thus, using (109b), we have

$$\psi_\lambda d\lambda = 8\pi kT \lambda^{-4} d\lambda \quad (110)$$

as the equation for the spectral energy distribution in black-body radiation as deduced from the theorem of the equipartition of energy. This is the famous Rayleigh-Jeans formula.

It is obvious, at once, however, that Eq. (110) cannot represent the distribution of energy *throughout the spectrum* of a black body. For at a given temperature T , the value of ψ_λ , according to this equation, should vary inversely as the fourth power of the wave length and should *increase rapidly to infinity as the wave length approaches zero*,

whereas the actual curves shown in Fig. 48 rise to a maximum with decreasing wave length *and then decrease to zero*. Only at long wave lengths, in fact, is the Rayleigh-Jeans expression found to be in accord with experiment. Thus it is clear that the fundamental principles upon which the deduction of the formula is based are only partly valid.

A little consideration shows, in fact, that the law of the equipartition of energy, as stated above, and the computation of the number of degrees of freedom in the electromagnetic field, resulting in Eq. (109b), cannot simultaneously be correct. For the field should be capable of containing waves of wave length varying from infinity to zero. Hence, by integration of Eq. (109b) from $\lambda = \infty$ to $\lambda = 0$, we find that the number of degrees of freedom in the field should be infinite. And, if the law of the equipartition of energy is correct, it follows that the energy density of black-body radiation, when in equilibrium with the walls of an enclosure at a given temperature, should be vastly greater, indeed infinitely greater, than the energy density in the walls themselves. In reality quite the opposite is the case. Thus the radiation density in an enclosure at 1000°K . is only of the order of 8×10^{-3} erg per cm^3 , whereas the energy density of the thermal agitation of the molecules of the walls, say of iron, is of the order of 10^{10} ergs per cm^3 .

89. Other Formulas for Black-body Radiation.—A formula quite different from that suggested by Rayleigh was proposed by Wien in 1896, on the basis of special assumptions as to the mechanism of absorption and emission of radiation by gas molecules. Wien's formula is

$$\psi_\lambda = c_1 \lambda^{-5} e^{-\frac{c_2}{\lambda T}}. \quad (111)$$

By adjusting the constants c_1 and c_2 , this formula can be made to fit the observations at *high* temperatures (*cf.* Fig. 56 below); but it fails definitely at *low* temperatures.

Finally Planck, in 1901,¹ by introducing a radical innovation quite at variance with previous concepts, discovered a function of λT which gave a formula in complete agreement with experiment. *This was the birth of the quantum theory*. We shall devote the remainder of this chapter to a discussion of the ideas leading up to Planck's quantum hypothesis and of his new radiation law.

Planck arrived at his formula as the result of a persistent effort to make the theory fit the facts. The experiments of Hertz on electromagnetic waves had seemed to give final confirmation to the electromagnetic theory of light, and this convinced Planck that the key to

¹ *Ann. d. Physik*, vol. 4, p. 553 (1901).

the black-body spectrum would be found in the laws governing the absorption and emission of radiation by electrical oscillators. We may imagine that the walls of an isothermal enclosure contain electrical oscillators of all frequencies, essentially similar to the Hertzian oscillator, and that the emission and absorption of radiation by the walls are caused by these oscillators.

Investigation by means of electromagnetic theory, however, led Planck, before 1900, to the conclusion that an electrical oscillator, in the long run, would affect only radiation of the same frequency as that of the oscillator itself. Thus, in the state of equilibrium, there would be a definite ratio between the density of radiation of any frequency ν and the average energy of the oscillators of that frequency. The problem of the black-body spectrum was thus reduced to the problem of the average energy of an oscillator at a given temperature. Now we have seen (Secs. 84 and 85) that according to classical theory a harmonic oscillator in thermal equilibrium with its surroundings would have an average energy equal to kT . If this conclusion from the theory is combined with the value obtained by Planck for the ratio between oscillator energy and radiation density in the enclosure, the result is again the Rayleigh-Jeans radiation law, which was derived above in a different manner. Planck himself, however, did not accept the principle of equipartition for the oscillators. On the basis of a different assumption he was led, at first, to Wien's law for the radiation density as stated in Eq. (111); and he believed for a time that this formula must necessarily be correct.

In 1900, however, new measurements of the black-body spectrum by Lummer and Pringsheim, and by Rubens and Kurlbaum, showed definitely that Wien's formula, if adjusted to fit the observations at high temperatures, gave values at low temperatures that were certainly too small. So Planck set to work to find an improved formula. First he discovered an empirical modification of Wien's formula that fitted the observations. Then he sought to modify the statistical theory so that it would lead to this new formula. He succeeded in doing this only after making a new assumption that broke drastically with classical principles.

It would take us too far afield to give a complete theoretical deduction of Planck's radiation law, but the most important ideas that are involved can easily be described.

90. Distribution of Harmonic Oscillators in Thermal Equilibrium.

In order to complete the theory of black-body radiation along the lines followed by Planck, we need to know the average energy of an oscillator at a given temperature. This is a problem in statistical

mechanics.¹ The problem is solved by finding first the distribution of energy among a large number of oscillators in thermal equilibrium. The characteristic feature of this distribution is well illustrated by the more familiar case of Maxwell's law for molecular velocities.

Maxwell's law can be stated as follows.² Let the components of the velocity of a molecule in the directions of the x -, y -, and z -axes be v_x , v_y , v_z , respectively. Among the N molecules in any chosen region (N being large), choose those which have an x -component of velocity lying between a given value v_x and $v_x + dv_x$, a y -component between v_y and $v_y + dv_y$, and a z -component between v_z and $v_z + dv_z$. Then the number of these molecules is

$$dN = NA e^{-mv^2/2kT} dv_x dv_y dv_z, \quad (112a)$$

where m is the mass of a molecule, k is Boltzmann's constant, T is the absolute temperature, and v is the molecular speed, so that

$$v^2 = v_x^2 + v_y^2 + v_z^2.$$

A is a constant, independent of v_x , v_y , and v_z , of such magnitude that $\int dN = N$, the total number of molecules.

The theoretical argument from statistical mechanics actually leads to a slightly different expression, however, and this is better suited to our purpose. The momentum is used to specify the motion instead of the velocity; the reason for this change, which will not be stated here, is closely related to the fact that, in a collision between 2 molecules, momentum is conserved, whereas velocity is not. The components of the momentum are

$$\begin{aligned} p_x &= mv_x, & p_y &= mv_y, & p_z &= mv_z. \\ \therefore dp_x &= m dv_x, & dp_y &= m dv_y, & dp_z &= m dv_z. \end{aligned}$$

Obviously the values of p_x , p_y , and p_z serve as well to specify the motion of a molecule as do those of v_x , v_y , and v_z . The kinetic energy

$$\zeta = \frac{1}{2} mv^2$$

is also introduced. In terms of these quantities, the above expression for dN can be written

$$dN = NC e^{-\zeta/kT} dp_x dp_y dp_z, \quad (112b)$$

where $C = A/m^3$. The distribution of the molecules among the various possible values of the momentum is thus determined by the "Boltzmann factor," $e^{-\zeta/kT}$.

¹ Cf. KENNARD, E. H., "Kinetic Theory of Gases," Chap. IX, 1938; TOLMAN, R. C., "Principles of Statistical Mechanics," 1938; MAYER, J. E. and M. G., "Statistical Mechanics," 1940.

² Cf. KENNARD, *op. cit.*, Sec. 28; LOEB, L. B., "Kinetic Theory of Gases."

For a set of harmonic oscillators, a very similar expression is obtained from statistical mechanics. We shall consider linear oscillators only, so that a single coordinate x suffices to specify the position of the vibrating mass m ; let the corresponding momentum be

$$p = m \frac{dx}{dt}.$$

In this case we have potential energy as well as kinetic energy to deal with. Let the potential energy be $\frac{1}{2}\beta x^2$; the kinetic energy can be written $\frac{1}{2}m(dx/dt)^2 = p^2/2m$. Thus the total energy is

$$\epsilon = \frac{p^2}{2m} + \frac{1}{2}\beta x^2. \quad (113)$$

Elementary theory gives for the frequency of vibration of such an oscillator

$$\nu = \frac{1}{2\pi} \sqrt{\frac{\beta}{m}}. \quad (114)$$

The statistical law for a set of N such oscillators in thermal equilibrium is, then, that at any given moment dN of them will have x and p lying in the ranges dx and dp , respectively, where

$$dN = NC e^{-\epsilon/kT} dx dp, \quad (115)$$

C being, like A above, a constant.

This result is more easily utilized for our present purpose, however, if it is thrown into a slightly different form.

Let us construct a plot on which x and p are taken as Cartesian coordinates (Fig.

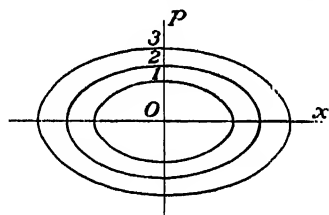


FIG. 55.—The momentum-coordinate plane for a harmonic oscillator.

55). On this plot $dx dp$ is an element of area. Now, as the oscillator moves, x and p change with the time in such a way that the energy ϵ remains constant; hence the point whose coordinates are x and p , representing the instantaneous state of the oscillator, moves on a curve given by Eq. (113) with a fixed value of ϵ . This curve is obviously an ellipse.

It is natural, therefore, to take as an element of area, instead of $dx dp$, the elliptical ring between two ellipses corresponding, respectively, to energies ϵ and $\epsilon + d\epsilon$. The area of such a ring can be expressed in terms of ϵ . For the entire area inside one of the ellipses is

$$S = \pi p_m x_m,$$

where p_m , x_m are its semiaxes, representing the maximum values of p and x , respectively, during a vibration. Putting in (113) first $p = p_m$ and $x = 0$, then $p = 0$ and $x = x_m$, we find

$$p_m = \sqrt{2m\epsilon}, \quad x_m = \sqrt{\frac{2\epsilon}{\beta}}.$$

Hence

$$S = 2\pi\epsilon \sqrt{\frac{m}{\beta}} = \frac{\epsilon}{\nu} \quad (116)$$

by (114). The area of a ring corresponding to $d\epsilon$ is thus

$$dS = \frac{d\epsilon}{\nu}.$$

Substituting this value of dS for $dx dp$ in (115) and writing C_1 for C/ν , we have

$$dN = NC_1 e^{-\epsilon/kT} d\epsilon \quad (117)$$

as the number of oscillators of which the energy lies between a given value ϵ and $\epsilon + d\epsilon$ when a large number N of them are in thermal equilibrium at absolute temperature T .

91. Average Energy of Oscillators in Equilibrium.—If the distribution law is known for any set of physical systems, their average energy is easily found by integration.

The method may be illustrated in the case of gases. If we multiply dN , as given by Eq. (112a), by the expression for the kinetic energy of a molecule of which the components of velocity are v_x , v_y , v_z , we have

$$\frac{1}{2} m (v_x^2 + v_y^2 + v_z^2) N A e^{-mv^2/2kT} dv_x dv_y dv_z,$$

where $v^2 = v_x^2 + v_y^2 + v_z^2$, as the total kinetic energy of the dN molecules. We now obtain the total kinetic energy K of all of the molecules if we integrate this expression for all values of v_x , v_y , and v_z from $-\infty$ to ∞ :

$$K = \frac{1}{2} N m A \iiint (v_x^2 + v_y^2 + v_z^2) e^{-mv^2/2kT} dv_x dv_y dv_z. \quad (118)$$

Now

$$\begin{aligned} & \iiint v_x^2 e^{-mv^2/2kT} dv_x dv_y dv_z \\ &= \left(\int_{-\infty}^{\infty} v_x^2 e^{-mv_x^2/2kT} dv_x \right) \left[\int_{-\infty}^{\infty} e^{-m(v_y^2+v_z^2)/2kT} dv_y dv_z \right]. \end{aligned}$$

But by an integration by parts¹

¹ In the formula $\int u dv = uv - \int v du$, put $u = kT v_x/m$, $dv = m v_x e^{-mv_x^2/2kT} dv_x/kT$.

$$\begin{aligned}\int_{-\infty}^{\infty} v_x^2 e^{-mv_x^2/2kT} dv_x &= -\frac{kT}{m} v_x e^{-mv_x^2/2kT} \Big|_{v_x=-\infty}^{v_x=\infty} + \frac{kT}{m} \int_{-\infty}^{\infty} e^{-mv_x^2/2kT} dv_x \\ &= \frac{kT}{m} \int_{-\infty}^{\infty} e^{-mv_x^2/2kT} dv_x,\end{aligned}$$

since in the integrated term $e^{-mv_x^2/2kT} \rightarrow 0$ as $v_x \rightarrow \pm \infty$ much faster than $|v_x| \rightarrow \infty$. Thus

$$\iiint v_x^2 e^{-mv^2/2kT} dv_x dv_y dv_z = \frac{kT}{m} \iiint e^{-mv^2/2kT} dv_x dv_y dv_z.$$

But, to make the whole number of molecules equal to N ,

$$N = \int dN = NA \iiint e^{-mv^2/2kT} dv_x dv_y dv_z,$$

from (112a). Eliminating the last integral between these two equations, we find

$$\iiint v_x^2 e^{-mv^2/2kT} dv_x dv_y dv_z = \frac{kT}{mA}.$$

The terms in v_y^2 and v_z^2 in (118) yield the same result. Hence, by (118),

$$K = \frac{1}{2} N m A \frac{3kT}{mA} = \frac{3}{2} N kT,$$

and the mean kinetic energy of a molecule is

$$\frac{K}{N} = \frac{3}{2} kT.$$

This agrees with Eq. (104) in Sec. 85.

In making, now, a similar calculation for the harmonic oscillator, a slight modification will be introduced in order to connect with Planck's line of thought. Let us draw ellipses, as in Fig. 55, in such a way as to divide the entire xp plane into elliptical rings of equal area. Let us number these rings 0, 1, 2, 3, \dots from the center outward. Then, if we denote the area of a ring by h , the total area inside ring number τ is

$$S = \tau h;$$

and by Eq. (116) the energy of an oscillator represented by a point on the inner boundary of this ring is

$$\epsilon = S\nu = \tau h\nu.$$

Thus the range of energy represented by points lying on a ring is

$$\Delta\epsilon = \nu\Delta S = h\nu,$$

since $\Delta\tau = 1$ from one ring boundary to the next.

The number of oscillators represented by points on ring number τ can be found by integrating dN as given by Eq. (117) over the range $\Delta\epsilon$ of ϵ within the ring. Calling this number N_τ , we thus find

$$N_\tau = NC_1 \int_{\Delta\epsilon} e^{-\epsilon/kT} d\epsilon.$$

If h and $\Delta\epsilon$ are small, however, we can treat $e^{-\epsilon/kT}$ as constant in this integral and equal to its value at a point on the inner boundary of the ring; then

$$\int_{\Delta\epsilon} e^{-\epsilon/kT} d\epsilon = e^{-\epsilon/kT} \int_{\Delta\epsilon} d\epsilon = e^{-\epsilon/kT} \Delta\epsilon.$$

Thus, replacing $NC_1 \Delta\epsilon$ by N_0 , we have for the number of oscillators on ring number τ

$$N_\tau = N_0 e^{-\frac{\tau h\nu}{kT}}. \quad (119)$$

Obviously N_0 represents the number of oscillators in ring number 0. Its connection with the total number of oscillators N is easily found by noting that

$$N = N_1 + N_2 + \dots = N_0(1 + e^{-\frac{h\nu}{kT}} + e^{-\frac{2h\nu}{kT}} + \dots) = \frac{N_0}{1 - e^{-h\nu/kT}}.$$

(For $1/(1-x) = 1 + x + x^2 + \dots$).

The *total energy* of the oscillators can then be found by multiplying the number in each ring by the energy of an oscillator when in that ring, for which, since h and $\Delta\epsilon$ have been assumed to be small, we can use the value $\tau h\nu$ belonging to the inner boundary, and then summing over all rings. Thus the total energy is

$$\begin{aligned} E &= (N_0 \times 0) + (N_0 e^{-\frac{h\nu}{kT}} \times h\nu) + (N_0 e^{-\frac{2h\nu}{kT}} \times 2h\nu) + \dots \\ &= N_0 h\nu e^{-h\nu/kT} (1 + 2e^{-h\nu/kT} + 3e^{-2h\nu/kT} + \dots), \\ &= \frac{N_0 h\nu e^{-h\nu/kT}}{(1 - e^{-h\nu/kT})^2}, \end{aligned}$$

since the last series is of the form

$$1 + 2x + 3x^2 + \dots = \frac{1}{(1-x)^2}.$$

Dividing this expression for E by that found above for N , the total number of oscillators, we thus find finally for the *average energy* of an oscillator

$$\bar{\epsilon} = \frac{E}{N} = \frac{h\nu e^{-h\nu/kT}}{1 - e^{-h\nu/kT}} = \frac{h\nu}{e^{h\nu/kT} - 1}. \quad (120a)$$

The *classical value* for $\bar{\epsilon}$ can now be obtained from this expression by letting $h \rightarrow 0$. This converts $\Delta\epsilon$ into an infinitesimal, $d\epsilon$, so that our sums become integrals, as in the preceding calculation for the mean energy of a molecule. We use the series

$$e^x = 1 + x + \frac{1}{2}x^2 + \dots,$$

so that to the first order in h

$$e^{h\nu/kT} - 1 = \frac{h\nu}{kT} \dots$$

Then it is clear that according to (120a), as $h \rightarrow 0$,

$$\bar{\epsilon} \rightarrow kT'.$$

This is the same value for the average energy of a harmonic oscillator that was deduced from the equipartition of energy in Secs. 84 and 85.

92. Planck's Quantum Hypothesis.—We are now in position to introduce the new assumption made by Planck, which initiated the quantum theory.¹ His assumption amounted to *keeping the quantity h finite* instead of letting it go to zero.

In the first formulation of the new theory, Planck assumed that the oscillators associated with a given ring all have the energy proper to the inner boundary of that ring. According to this assumption, the energy of an oscillator *cannot vary continuously but must take on one of the discrete set of values*: $0, h\nu, 2h\nu, \dots, \tau h\nu, \dots$. Actually, Planck's original assumption was somewhat more general. He assumed that the energy of the oscillator must always be an integral multiple of a certain quantity, ϵ_0 . He then showed, however, that ϵ_0 must be proportional to ν if the radiation law is to harmonize with the Wien displacement law. Thus he assumed that $\epsilon_0 = h\nu$, where h is a constant of proportionality. The connection between the constant h and areas on the xp plane for the oscillator, as described above, was recognized by Planck later.

It must be emphasized that the assumption of a *discrete set* of possible energy values, or *energy levels*, for an oscillator was completely at variance with classical ideas. According to this assumption, if the energies of a large number of oscillators were measured, some might be found to have zero energy; some, energy $h\nu$ ergs each; others, $2h\nu$; and so on. But not a single oscillator would be found which had energy, say $1.75 h\nu$ or $3.94 h\nu$. When the energy of a given oscillator changes, therefore, it must change suddenly and discontinuously. According to the older conceptions, on the other hand,

¹ PLANCK, *Ann. d. Physik*, vol. 4, p. 553 (1901).

the interchange of energy between two "systems," as, for example, between one gas molecule and another or between radiation and oscillators, should be a perfectly *continuous* process. Thus, if one could follow the "history" of a particular gas molecule or oscillator and observe its energy at short intervals of time over a very long period, one would find, of course, that its energy varied over enormously wide limits. But, if the numerical values of the energy observed at the various instants were to be arranged in a sequence as to magnitude, this sequence would become more and more *continuous* as the number of observations becomes greater. Or, if, instead of observing a single molecule over a long period, one were to observe, at a given instant, the energy of a very large number of gas molecules or of oscillators chosen at random, one ought to find that a similar sequence of values of energy would approach more and more nearly to continuity as the number of observations becomes larger. This continuity of these energy values is not only in accord with but is also imperatively demanded by classical physics. For example, the electric and magnetic vectors in a light wave may have any values whatsoever, *from zero up*; and, accordingly, the wave may have any intensity, *from zero up*. Furthermore, the emission and absorption of this energy by the walls of an enclosure should, likewise, be a perfectly continuous process.

The problem of the absorption and emission of radiation, in fact, presented serious difficulties for the new theory. If the energy of an oscillator can vary only discontinuously, the *absorption and emission of radiation* must likewise be *discontinuous processes*. As long as the oscillator remains in one of its "quantum states," as we now call them, with its energy equal to one of the allowed discrete values, it cannot be emitting or absorbing radiation according to the laws of classical physics, for then the conservation of energy would be violated. This is entirely contrary to classical electromagnetic theory; for we have seen (Sec. 38) that classical theory absolutely requires an isolated, accelerated electrical charge to radiate energy.

According to Planck's new theory, *emission* of radiation could occur only when the oscillator "jumps" from one energy level to another; if it jumps down to the next lower energy level, the energy $h\nu$ that it loses is emitted in the form of a short pulse of radiation. *Absorption* was also assumed at first to be discontinuous. An oscillator, Planck assumed, can absorb a quantum $h\nu$ of radiant energy and jump instantaneously (or nearly so) up to its next higher energy level. This assumption met with difficulties, however. For the quantum of radiant energy emitted by an oscillator, according to the classical

wave theory, would spread out over an ever expanding wave front, and it is hard to see how another oscillator could ever gather this energy together again so as to absorb it all and thereby acquire the energy for an upward quantum jump. Absorption ought, therefore, on Planck's theory, to be impossible.

For this reason, Planck later modified his theory so as to allow the oscillators to absorb in a continuous manner, only the process of emission being discontinuous. The energy of an oscillator could then take on all values, as in classical theory; but, every time that the energy passed one of the critical values, $\tau h\nu$, there was assumed to exist a certain chance that the oscillator would jump down to a lower energy level, emitting its excess energy as a quantum of radiation. This came to be known as the "second form" of Planck's quantum theory. Calculation showed that, on the theory as thus modified, the average energy of an oscillator, when in thermal equilibrium with its surroundings, would be

$$\bar{\epsilon} = \frac{h\nu}{e^{h\nu/kT} - 1} + \frac{1}{2}h\nu, \quad (120b)$$

in place of the value given in Eq. (120a) above.

Having read thus far, the student may perhaps have reached a state of fairly complete confusion as to *what were the essential* assumptions of Planck's quantum theory! This confusion can be no worse than that which existed in the minds of most physicists in the year, say, 1911. The situation was made still more puzzling by the success of Einstein's theory of the photoelectric effect, described above (Sec. 49); for Einstein assumed not only that radiation came in quantized spurts but that each spurt was closely concentrated in space, contrary to the wave theory. One of the principal aims of the rest of this book will be to show how the theory has gradually become clarified. To assist the student at this point, it may be stated that the following two new ideas introduced by Planck have been retained permanently and form a part of modern "wave mechanics" (Chap. VII):

1. *An oscillator, or any other physical system, has (in general) a discrete set of possible energy values or levels; energies intermediate between these allowed values never occur. [In wave mechanics, however, the energy of level number τ for a harmonic oscillator is $(\tau + \frac{1}{2})h\nu$, and their average energy at temperature T is given by (120b), not (120a).]*

2. *The emission and absorption of radiation are associated with transitions or jumps between two of these levels, the energy thereby lost*

or gained by the oscillator being emitted or absorbed, respectively, as a quantum of radiant energy, of magnitude $h\nu$, ν being the frequency of the radiation.

It should be emphasized that Planck's revolutionary assumptions were not based upon an extension of the ordinary lines of reasoning of classical physics. Quite the contrary: they represented an *empirical modification* of classical ideas made in order to bring the theoretical deductions into harmony with experiment.

Had the magnitude of the quantum of energy turned out to be not $h\nu$ but something *independent of the frequency*, the new theory might well have taken the form of a simple atomicity of energy, similar to the atomicity of electricity represented by the electronic charge. Such is not the case, however. Rather, it is the new universal constant h that represents the essentially new element introduced into physics by the quantum theory. We shall find h playing an important part in a wide variety of atomic phenomena. It may be said, however, even today, that we know a great deal more about the *numerical value* of h than we do of its *physical significance*.

93. Planck's Radiation Law. (a) *The New Radiation Formula.*—Planck derived his new radiation formula by considering the interaction between the radiation inside an isothermal enclosure and electrical oscillators which he imagined to exist in the walls of the enclosure. A modern version of this deduction, in which gaseous atoms take the place of oscillators, first given by Einstein, will be sketched later [Sec. 125(b)]. The most satisfactory procedure, however, seems to be to combine Planck's expression for the mean energy of an oscillator with the analysis of the electromagnetic field by the method of Rayleigh and Jeans, in which the various modes of oscillation of the field inside an enclosure are treated as if they were oscillators.

In Sec. 87, Eq. (109b), we found that there would be

$$8\pi \frac{d\lambda}{\lambda^4}$$

such modes of oscillation or degrees of freedom per unit volume in the wave length range λ to $\lambda + d\lambda$. If we multiply this number, not by kT , the mean energy of an oscillator according to the law of equipartition, but by $\bar{\epsilon}$ as given by Eq. (120a) above, we obtain

$$8\pi \frac{d\lambda}{\lambda^4} \frac{h\nu}{e^{h\nu/kT} - 1}.$$

This is the value of $\psi_\lambda d\lambda$, the energy density belonging to the range $d\lambda$. Let us substitute in it $\nu = c/\lambda$, c being the speed of light. Thus we

obtain, as Planck's new radiation law,

$$\psi_{\lambda} = \frac{8\pi ch}{\lambda^5} \frac{1}{e^{ch/\lambda kT} - 1}. \quad (121)$$

The emissivity of a black body, e_{λ} , is then equal to ψ_{λ} as given by this formula multiplied by $c/4$ [cf. Eq. (83b) in Sec. 75]. [Strictly speaking, we should have used for $\bar{\epsilon}$ the value given by Eq. (120b), which agrees with the value obtained from wave mechanics. The effect of this change would be to add in ψ_{λ} a term independent of temperature, $4\pi h\nu/\lambda^4$. Since only *changes* in ψ are perceptible, however, this term would be without physical effect. Its true significance is still unknown, and it is never included in the expression for ψ_{λ} or e_{λ} .]

(b) *Comparison of Planck's Formula with Experiment.*—If Planck's formula is to fit the observations, it must (1) approximate to Wien's formula at short wave lengths (or, at any wave length provided the temperature is low enough) and (2) approximate to the Rayleigh-Jeans formula at long wave lengths or high temperatures. That it approximates to Wien's formula is obvious. For if the value of the product λT is small enough, the exponential term in the denominator in (121) is very much larger than unity, so that the -1 can be dropped, and (121) then becomes

$$\psi_{\lambda} = 8\pi ch\lambda^{-5} e^{-ch/\lambda kT};$$

this agrees with Eq. (111) in Sec. 89 if in that equation we make $c_1 = 8\pi ch$, $c_2 = ch/k$. For *large* values of λT , on the other hand, we may expand the denominator in (121) by means of the series,

$$e^x = 1 + x + \frac{x^2}{2} + \cdots,$$

obtaining

$$e^{\frac{ch}{\lambda kT}} - 1 = \frac{ch}{\lambda kT} + \frac{c^2 h^2}{2\lambda^2 k^2 T^2} + \cdots$$

If λT is large enough, only the first term of this series need be kept; then we have, from (121),

$$\psi_{\lambda} = \frac{8\pi kT}{\lambda^4},$$

approximately. This agrees with the Rayleigh-Jeans formula as stated in Eq. (110) in Sec. 88.

In Fig. 56 is shown a comparison between the several spectral-energy distribution formulas and the experimental data. The circles show observations by Coblentz¹ on the energy distribution in

¹ *Bureau of Standards, Bull.*, vol. 13, p. 476 (1916).

the spectrum of a black body at 1600°K. The full line shows the distribution predicted by Planck's formula. The lower dotted line, which coincides with the full line from short wave lengths up to about $\lambda = 2.2\mu$, corresponds to Wien's formula as given in Eq. (111). The upper dotted line is from the Rayleigh-Jeans formula. The superiority of Planck's formula is at once evident. Thus, whatever one may think of the theoretical assumptions and reasoning by means of which Planck arrived at his formula, there seems to be no doubt that it correctly represents the observations on black-body radiation.

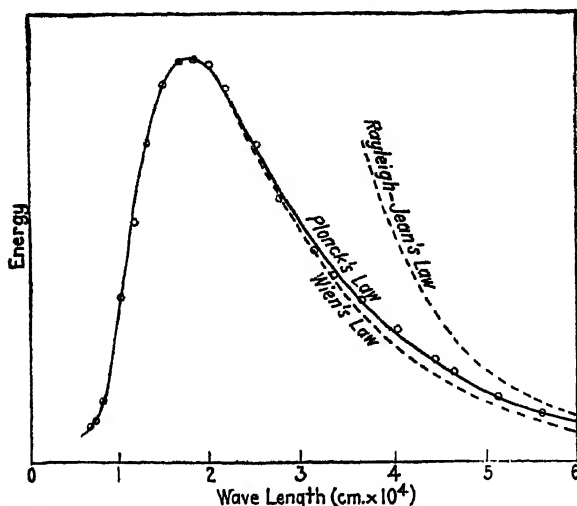


FIG. 56.—Comparison of the three radiation laws with experiment at 1600°K. The ordinate represents ψ_λ or e_λ , on an arbitrary scale.

The reason that the curve for Planck's formula drops below the Rayleigh-Jeans formula is to be found, of course, in the failure of the classical principle of the equipartition of energy. At short wave lengths, the -1 can be dropped in the denominator in Eq. (120b), and we can write, as the approximate average energy of an oscillator,

$$\bar{\epsilon} = h\nu e^{-h\nu/kT} + \frac{1}{2} h\nu.$$

Thus, the part of $\bar{\epsilon}$ that varies with temperature decreases rapidly toward zero as $\nu \rightarrow \infty$. The high-frequency modes of oscillation of the electromagnetic field in the enclosure, which should all have mean energy kT according to classical theory, remain, therefore, almost entirely in their lowest quantum states and so contribute nothing to the observable density of radiant energy.

(c) *Planck's Formula and the Thermodynamic Theory.*—In Sec. 83 we arrived, by means of rigorous arguments based on thermodynamical

reasoning, at the conclusion that the correct formula for the emissive power of a black body should have the form given in Eq. (103) in that section, or

$$e_{\lambda} = \frac{F(\lambda T)}{\lambda^5}.$$

Since $e_{\lambda} = c\psi_{\lambda}/4$, according to Eq. (83b) in Sec. 75, Planck's formula can be written, from (121),

$$e_{\lambda} = \frac{F(\lambda T)}{\lambda^5}, \quad F(\lambda T) = \frac{2\pi c^2 h}{e^{\frac{ch}{\lambda k T}} - 1}.$$

Thus Planck's formula has the required form.

We may be sure, therefore, that Planck's formula will lead to the Wien displacement law and to the Stefan-Boltzmann law, which were deduced from the thermodynamical reasoning (Secs. 78, 82). It may be instructive, however, to verify these conclusions directly, at least in part.

A special feature resulting from Wien's displacement law is that

$$\lambda_m T = \text{const.},$$

where λ_m is the wave length at which ψ_{λ} and e_{λ} have their maximum values. Let us find the value of λ_m as given by Planck's formula. To do this, we set $d\psi_{\lambda}/d\lambda = 0$, using Eq. (121):

$$\frac{d\psi_{\lambda}}{d\lambda} = -\frac{40\pi ch}{\lambda^6 (e^{ch/\lambda k T} - 1)} + \frac{8\pi ch}{\lambda^5} \frac{ch}{\lambda^2 k T} \frac{e^{ch/\lambda k T}}{(e^{ch/\lambda k T} - 1)^2} = 0;$$

or, after simplification,

$$\left(1 - \frac{ch}{5\lambda k T}\right) e^{ch/\lambda k T} = 1.$$

This last equation can be written

$$\left(1 - \frac{x}{5}\right) e^x = 1, \quad x = \frac{ch}{\lambda k T}.$$

One root is obviously $x = 0$, but this one does not interest us. Any other positive root must be < 5 , for, if $x > 5$, the factor in parentheses is negative. If $x = 5$, the left-hand member is zero, whereas, if $x = 4$, it is much bigger than 1. Another root must lie, therefore, just below 5. To facilitate finding it, let us write $5 - x = y$, so $x = 5 - y$ and the equation becomes

$$y e^{5-y} = 5$$

or

$$\log_e y - y = \log_e 5 - 5 = -3.3906.$$

To make the solution still easier, set $y = z/10$; then

$$-\log_e z + \frac{z}{10} = 3.3906 - \log_e 10 = 1.0880.$$

It is now easy to see, from a table of natural logarithms, that to three figures $z = 0.349$; hence, $y = 0.0349$, $x = 4.965$. Thus, λ_m is given by

$$\frac{ch}{\lambda_m k T} = 4.965$$

or

$$\lambda_m T = \frac{ch}{4.965 k}. \quad (122)$$

Thus, $\lambda_m T$ is equal to a constant, independent of T . In this respect, therefore, Planck's formula is in agreement with the Wien displacement law.

To compare Planck's formula with the Stefan-Boltzmann law of total radiation, it is necessary to integrate $\psi_\lambda d\lambda$ over the range of wave lengths from $\lambda = 0$ to $\lambda = \infty$. We thus find for the total energy in unit volume, from (121),

$$\psi = \int_0^\infty \psi_\lambda d\lambda = 8\pi ch \int_0^\infty \frac{d\lambda}{\lambda^5} \frac{1}{e^{ch/\lambda k T} - 1}.$$

Let us change the variable here to

$$x = \frac{ch}{\lambda k T}; \quad \therefore \lambda = \frac{ch}{x k T}, \quad d\lambda = -\frac{ch dx}{x^2 k T}.$$

Then

$$\psi = \frac{8\pi k^4 T^4}{c^3 h^3} \int_0^\infty \frac{x^3 dx}{e^x - 1}. \quad (123)$$

The integral here represents a definite number, independent of T . Hence, the equation shows at once that ψ is proportional to the fourth power of the absolute temperature, which is one way of stating the Stefan-Boltzmann law.

By means of Eqs. (122) and (123), it is possible to calculate values of k and h from measurements of ψ and of $\lambda_m T$, the value of c being

already known. The integral in (123) can be evaluated;¹ it is equal to $\pi^4/15$. Hence, Eq. (123) can be written

$$\psi = aT^4, \quad a = \frac{8 \pi^5 k^4}{15 c^3 h^3}. \quad (124a,b)$$

In his original paper (1901), Planck used the values

$$a = 7.061 \times 10^{-15},$$

$\lambda_m T = 0.294$ and found, from Eqs. (122) and (124b),

$$h = 6.55 \times 10^{-27}, \quad k = 1.346 \times 10^{-16}.$$

This was the first calculation of h and the best up to that time of k . At the present time, the best experimental values of the constant a , or ψ/T^4 , and of $\lambda_m T$, are probably²

$$a = 7.652 \times 10^{-15} \text{ erg cm.}^{-3} \text{ deg.}^{-4}, \quad \lambda_m T = \frac{1.432}{4.9651} = 0.2884 \text{ cm. deg.}$$

If we insert these values of a and of $\lambda_m T$ into Eqs. (122) and (124b), also $c = 2.9980 \times 10^{10}$, and then solve for h and k , we find

$$h = 6.576 \times 10^{-27}, \quad k = 1.377 \times 10^{-16} \text{ erg deg.}^{-1}$$

Nowadays, however, the most accurate procedure is to calculate k from the values of the electronic charge e and of the faraday (Sec. 45), the result being $k = 1.381 \times 10^{-16}$. If this value of k is inserted in Eqs. (124b) and (122) together with the values just given for a , $\lambda_m T$, and c , the two values 6.603×10^{-27} , 6.596×10^{-27} are obtained for h . These numbers agree fairly well with values of h obtained by other methods, such as the photoelectric effect. As the *present best values* of h and k may be taken³

$$h = 6.61 \times 10^{-27} \text{ erg sec.}; \quad k = 1.381 \times 10^{-16} \text{ erg deg.}^{-1}$$

¹ Since $(e^x - 1)^{-1} = e^{-x}(1 - e^{-x})^{-1} = e^{-x} + e^{-2x} + e^{-3x} + \dots$, and

$$\int_0^\infty x^n e^{-nx} dx = \frac{n!}{n^{n+1}},$$

we have

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = 6 \left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \dots \right).$$

The value of the series written here in parentheses, as given on p. 140 of the "Smithsonian Mathematical Formulæ and Tables of Elliptic Functions" (Smithsonian Institution, 1922) is $\pi^4/90$. Cf. also Bierens de Hahn, "Nouvelle Table d'Intégrales Définies," No. 4, p. 124, where $B_{2n-1} = \frac{1}{2} \zeta_n$.

² Cf. BIRGE, *Rev. Modern Physics*, vol. 1, p. 1 (1929). Values of a are by Kussman (1924) and Coblenz (1920), of $\lambda_m T$, by Coblenz (1922) and Ladenburg.

³ The value of h is adopted according to the recommendation of Dunnington, *Rev. Modern Physics*, vol. 11, p. 65 (1939).

CHAPTER VI

THE NUCLEAR ATOM AND THE ORIGIN OF SPECTRAL LINES

We have seen in the last chapter that attempts to explain the experimentally observed laws of the distribution of energy in the continuous spectrum emitted by a black body were unsuccessful until Planck introduced the revolutionary concept of radiation quanta. Planck's hypothesis became both *possible and necessary*, because the very careful experiments of Lummer and Pringsheim had proved that Wien's law of temperature radiation was untenable.

A somewhat similar sequence of events is to be found in the development of our present extensive knowledge of the *line* spectra of atoms and molecules. Corresponding to the empirical laws of temperature radiation, there was accumulated a vast array of very accurate measurements of the wave lengths of lines in the spectra of various substances, considerable impetus being given to this work because of the rigorous demands of spectroscopy as a method of chemical analysis. From these data, certain relations were discovered empirically between the frequencies of various lines in the spectra of certain elements. These relations pointed to some fundamental mechanism, common to all atoms, as the origin of characteristic line spectra. The gradual accumulation of evidence bearing on the problem of atomic structure, on the one hand, and the increasing importance of these spectral relations, on the other, culminated, about 1913, in the proposal, by Bohr, of the theory of atomic structure and the origin of spectra which bears his name. This theory constituted an extension in a new field of the quantum theory which had been introduced by Planck to explain the law of the distribution of energy in temperature radiation.

In the present chapter, we shall begin by considering very briefly the development of the empirical laws of spectral series. Next we shall take up the lines of evidence that were accumulated bearing on the structure of the atom, culminating in the famous nuclear theory of atomic structure proposed by Rutherford. Then we shall consider the way in which these two lines of development were combined by Bohr in his famous quantum theory of the atom. Bohr's theory has been superseded in many respects by modern quantum

mechanics (or wave mechanics), but certain features of his theory retain a permanent interest.

94. Spectroscopic Units.—In spectral measurements we are concerned, sometimes with the *wave lengths* of lines, sometimes with *frequencies*. In stating wave lengths, which are always very small, various submultiples of the meter are employed as units of length in different parts of the spectrum, *viz.*,

The micron, symbol μ = 10^{-4} cm. (or 10^{-6} meter)

The millimicron, symbol $m\mu$ = 10^{-7} cm.

The angstrom,¹ symbol \AA = 10^{-8} cm.

(or Angström, symbol \AA .)

The X-unit, symbol X.U. = 10^{-11} cm.

The last named is used in the X-ray region of the spectrum.

Wave lengths of lines are of fundamental importance in the technique of spectroscopy and in spectroscopic analysis. In physical theory, on the contrary, *frequency* ν is more fundamental than wave length. We do not, however, measure frequencies *directly*; laboratory measurements yield *wave lengths*, and frequencies are *computed* from these and from the velocity of light c by the relation, $c = \nu\lambda$.

Frequencies may be expressed in vibrations per second, but this involves very large numbers. Furthermore, the calculated frequency is affected, not only by errors in the measured wave lengths, but also by any error there may be in the assumed velocity of light. For these reasons, it is customary in spectroscopy to use, instead of the frequency itself, the *wave number*, or number of waves per centimeter in vacuum, which we shall denote by $\bar{\nu}$. The unit for $\bar{\nu}$ is written cm.^{-1} ; it may be read "waves per centimeter." Thus

¹ Strictly speaking, the angstrom is not defined from the meter as a *primary* standard of length. Michelson and Benoist in 1895 and, later (1907), Fabry, Perot, and Benoist measured the wave length of the red cadmium line in terms of the standard meter. The two measurements were almost exactly in agreement, the wave length according to the latter measurement being

$$6,438.4696 \text{ angstroms.}$$

The International Union for Solar Research, in 1907, adopted this value of the wave length of the red cadmium line as the *primary* standard of wave lengths on the basis of which all other wave lengths were to be expressed. It is specified that the medium is to be dry air at 15°C . (hydrogen scale) and a pressure equal to 760 mm. Hg at a place where the acceleration due to gravity is $980.67 \text{ cm. sec.}^{-2}$. Formally, this amounts to a new definition of the angstrom in terms of the wave length of the cadmium line such that this wave length is *exactly* 6,438.4696 angstroms.

$$\tilde{\nu} = \frac{1}{\lambda} \text{ cm.}^{-1}, \quad \nu = c\tilde{\nu},$$

λ being the wave length in vacuum expressed in centimeters. Or, if λ_μ is the wave length in microns, or λ_A the wave length in angstroms,

$$\tilde{\nu} = 10^4 \times \frac{1}{\lambda_\mu} \text{ cm.}^{-1} = 10^8 \times \frac{1}{\lambda_A} \text{ cm.}^{-1}.$$

For example, the wave number of the red cadmium line

($\lambda = 6,438.4696 \text{ A}$, $= 0.6438\mu = 6.438 \times 10^{-5} \text{ cm.}$, approximately)

is $\tilde{\nu} = 15,531.64 \text{ cm.}^{-1}$; its frequency is 2.99796×10^{10} times this or $4.65632 \times 10^{14} \text{ sec.}^{-1}$. Very roughly, the visible region (4,000 to 7,000 A) extends from 14,000 to 25,000 cm.^{-1} ; from 14,000 down to 1,000 is the near infrared region (out to 10μ); from 1,000 to 100, the far infrared. The ultraviolet region extends from 25,000 to 100,000 cm.^{-1} (1,000 A). The principal X-ray region is from 10^7 to 10^8 cm.^{-1} .

The wave length of a line *in air* is slightly different from the wave length *in vacuum*. Since ordinary spectroscopic work is done in air, wave lengths above (about) 2,000 angstroms are commonly given *in air*.¹ Below 2,000 A, however, they are usually given in vacuum, since vacuum spectrographs are commonly employed for work in that region. The two wave lengths are related by the equation

$$\lambda_{vac} = \mu \lambda_{air},$$

where μ is the refractive index of air; thus the relation actually used ordinarily in obtaining wave numbers from optical wave lengths in air, or vice versa, is

$$\tilde{\nu} = \frac{1}{\mu \lambda_{air}}.$$

The difference between λ_{vac} and λ_{air} is somewhat less than 1 part in 3,000. Tables have been published to facilitate the conversion from λ_{air} to $\tilde{\nu}$ or vice versa.²

As a primary standard, it is customary to employ the wave length of the red cadmium line, referred to above. This line is very narrow and hence is easily measured with accuracy. A group of *secondary* standard lines based on this primary standard and extending at conveniently spaced intervals throughout the spectrum has also been adopted, and from these, in turn, a group of more closely spaced tertiary standards has been prepared. Measurements of wave

¹ Cf. the recent extensive table, "Massachusetts Institute of Technology Wave-length Tables," John Wiley & Sons, Inc., New York, 1939.

² Cf. KAYSER, H., "Tabelle der Schwingungszahlen," S. Hirzel, Leipzig, 1925.

lengths in the visible and near-visible region of the spectrum are now possible with an accuracy of at least 1 part in 1 million.

95. Early Search for Regularities in Spectra.—One of the first features of spectra to be noticed was that the spectrum of a given element depended a great deal upon its mode of excitation. When the spectrum was produced by a *spark*, many lines were observed which were absent, or at least much weaker, than when a steady arc or a flame was employed. These lines are said to form the *spark spectrum* (formerly, the “enhanced” spectrum) of the element, those lines present in the arc forming its *arc spectrum*. It was later established that arc spectra are emitted by neutral atoms, spark spectra by ionized atoms.

As soon as dependable wave-length measurements became available, numerous investigators, reasoning from the analogy of overtones

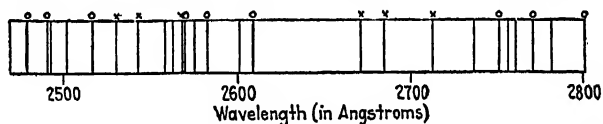


FIG. 57.—Some lines in the spectrum of zinc in a limited wave-length range in the ultraviolet. The lines marked with a cross (X) belong to one series of triplets, lines marked (O) belong to another series.

in acoustics, sought for harmonic relations in the lines found in the spectrum of a given element. This search proved fruitless; but certain relations of a different type were discovered.

Livinge and Dewar,¹ in the early eighties, emphasized the physical similarities occurring in the spectra of such elements as the alkali metals. They called attention to the successive *pairs* of lines in the arc spectrum of sodium and pointed out that these pairs were alternately “sharp” and “diffuse” and that they were more closely crowded together toward the short-wave-length end of the spectrum, suggesting some kind of series relation, which, however, they were unable to discover. A little later, Hartley² discovered an important numerical relationship between the components of doublets or triplets in the spectrum of a given element. If frequencies, instead of wave lengths, are used, Hartley found that *the difference in frequency between the components of a multiplet (i.e., doublet or triplet) in a particular spectrum is the same for all similar multiplets of lines in that spectrum.*

Hartley's law made it possible to isolate from the large number of lines in any given spectrum those groups of lines which were undoubtedly related. But the task was not an easy one. Figure 57 shows

¹ *Roy. Soc., Proc.*, vol. 29, p. 398 (1879); vol. 30, p. 93 (1880).

² *Chem. Soc. J.*, vol. 43, p. 390 (1883).

the lines in the zinc spectrum from about 2,500 to 2,800 Å. It is by no means obvious which of these many lines are associated by Hartley's law. Two triplets having constant frequency differences are marked by crosses (X) in Fig. 57. This same spectral region, however, contains another overlapping series of triplets, which are designated by circles (O). The sorting out of these related lines required a great deal of diligent and patient study.

The beginning of our knowledge of spectral-series formulas dates from the discovery by Balmer,¹ in 1885, that the wave lengths of the nine then known lines in the spectrum of hydrogen could be expressed by the very simple formula

$$\lambda = b \frac{n^2}{n^2 - 4}, \quad (125)$$

where b is a constant the numerical value of which, to give λ in angstroms, is 3,645.6 and n is a variable integer which takes on the successive values 3, 4, 5, . . . for, respectively, the first (beginning at the red), second, third, . . . line in the spectrum. Balmer compared the predictions of this formula with the best values then available for the wave lengths of the hydrogen lines. The four lines in the visible region had been measured by Ångström and others; five ultraviolet lines in the spectrum of white stars had been measured by Huggins. Table I, taken from Balmer's paper, shows the comparison of the formula with the measured wave lengths. The agreement is

TABLE I.—WAVE LENGTHS OF THE FIRST NINE HYDROGEN LINES
COMPUTED BY BALMER FROM HIS FORMULA $\lambda = 3,645.6 \frac{n^2}{n^2 - 4}$ ($a = 2$)

Line	n	λ (computed), angstroms	λ (observed), angstroms
H _{α}	3	6,562.08	6,562.10 (Ångström)
H _{β}	4	4,860.80	4,860.74 (Ångström)
H _{γ}	5	4,340.0	4,340.10 (Ångström)
H _{δ}	6	4,101.3	4,101.2 (Ångström)
H _{ϵ}	7	3,969.7	3,968.1 (Huggins)
H _{ζ}	8	3,888.6	3,887.5 (Huggins)
H _{η}	9	3,835.0	3,834.0 (Huggins)
H _{θ}	10	3,797.5	3,795.0 (Huggins)
H _{ι}	11	3,770.2	3,767.5 (Huggins)

seen to be excellent in the visible spectrum. The discrepancy between Balmer's computed values and the measurements of Huggins increases,

¹ *Ann. d. Physik*, vol. 25, p. 80 (1885).

however, to nearly 1 part in 1,000 for H_α . Balmer questioned whether this discrepancy indicated that the formula was only an approximation or whether the data were in error. Recent measurements have considerably revised Huggins' data but have also revealed the need for a slight correction to Balmer's formula [cf. Sec. 149(b)].

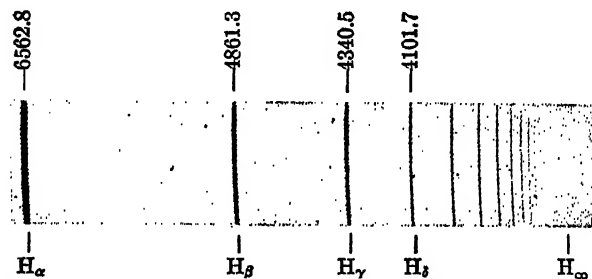


FIG. 58.—The Balmer series of atomic hydrogen, in emission. H_α marks the theoretical position of the series limit. (Photograph by G. Herzberg, *Ann. d. Physik*, vol. 84, p. 505 (1927); reprinted by courtesy of Prentice-Hall, Inc., New York, from G. Herzberg, "Atomic Spectra and Atomic Structure," 1927.)

Balmer correctly predicted that in this series of lines in hydrogen no lines of longer wave length than H_α would be found and that the series should "converge" at $\lambda = 3,645.6$ Å, since the fraction $\frac{n^2 - 4}{n^2}$ approaches unity as n becomes larger. In Fig. 58 is reproduced a photograph of the first few lines of the Balmer Series; in Fig. 59 is shown another photograph of the same series starting from the seventh line and showing the continuum



FIG. 59.—Higher members of the Balmer series, in emission, starting from the seventh line and showing the continuum. (After G. Herzberg, see credit under Fig. 58.)

line, more strongly exposed so as to bring out more lines. The theoretical limit of the series is marked in Fig. 58 by H_∞ .

The impetus which Balmer's discovery gave to work in spectral series is another illustration of the highly convincing nature of relations which are expressible in *quantitative* form. Soon after the publication of Balmer's work, intensive investigations in spectral series were initiated by Kayser and Runge and by Rydberg. We shall discuss here some of the results obtained by Rydberg in order to complete the background for Bohr's theory of atomic spectra.

96. Spectral Series and Their Interrelations.—In announcing his discovery, Balmer raised the question as to whether or not his formula might be a special case of a more general formula applicable to other series of lines in other elements. Rydberg set out to find such a formula.¹ Using the comparatively large mass of wave-length data then available and starting from the above-mentioned work of Liveing and Dewar, Rydberg isolated still other series of doublets and triplets of constant frequency difference, according to Hartley's law of constant wave-number separation. In all cases, these series showed a tendency to converge to some limit in the ultraviolet. He found that he could distinguish two types of such series: a type in which the lines were comparatively sharp, which he called, therefore, *sharp* series; and a type which, because the lines were comparatively broad,

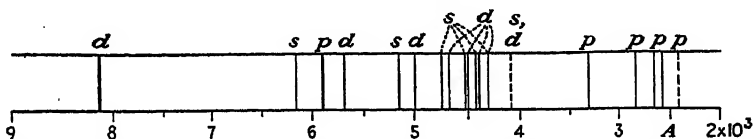


FIG. 60.—Principal lines in the spectrum of the neutral sodium atom: *p* = principal, *s* = sharp, *d* = diffuse series. Dotted lines indicate series limits. The first line of the sharp series, in the infrared at 11,393 Å, is not shown.

he called *diffuse* series. Both types of series occurred in the arc spectrum of the same element. In many arc spectra, he found also a third type of series in which the doublet or triplet spacing decreased as the frequency or ordinal number of the line increased, as if tending to vanish at the convergence limit of the series; these he called *principal* series, because they commonly contained the brightest and most persistent lines in the spectrum. The principal, sharp, and diffuse series of sodium are plotted in Fig. 60 and marked *p*, *s*, *d*, respectively, the first line of the sharp series at 11,393 Å not being shown.

In searching for a formula, Rydberg plotted the wave numbers of the lines, or their differences, in various ways. By a "line" in this connection is meant a complex of actual lines forming a doublet or triplet, which would be seen as a single line in a spectroscope of moderate resolving power. He observed that a similarity existed not only among the several series of a given element but also between series of different elements. He then showed that this similarity was *quantitative*. For example, in Fig. 61, the *differences* in wave numbers between successive lines in the sharp series of sodium, and similar differences for the diffuse series of the same element, are plotted as

¹ A brief account of Rydberg's work is given by him in *Phil. Mag.*, vol. 29, p. 331 (1890). For a fuller account see E. C. Baly's "Spectroscopy."

ordinates, represented by the circles, against the *ordinal number* of the difference as abscissas. The resolving power is supposed to be limited so that each doublet appears as a single line. It is found that exactly parallel curves can be drawn through the observed points; the two curves would coincide if the one for the diffuse series were displaced toward the right by a certain distance δ , equal to the length of the line aa or bb . Thus, we can say that the values of the successive differences for the sharp series are the same function of $(m + \delta)$ as the corresponding differences for the diffuse series are of m , where m is the ordinal number of the difference. Let us write $\tilde{\nu}_s$, $\tilde{\nu}_d$ for the wave numbers of sharp and diffuse lines, respectively, and $\Delta\tilde{\nu}_d$, $\Delta\tilde{\nu}_s$ for the successive differences. Then, if

$$\Delta\tilde{\nu}_d = f(m), \quad \Delta\tilde{\nu}_s = f(m + \delta). \quad (126a, b)$$

A similar relationship must exist, then, for $\tilde{\nu}_d$ and $\tilde{\nu}_s$ themselves. Let us write for the m th line of each series

$$\tilde{\nu}_{dm} = F_d(m), \quad \tilde{\nu}_{sm} = F_s(m). \quad (126c, d)$$

Then, from (126c and (126a),

$$\tilde{\nu}_{d,m+1} - \tilde{\nu}_{dm} = F_d(m + 1) - F_d(m) = f(m), \quad (126e)$$

provided the ordinal numbers of the differences are properly chosen. Similarly, for the sharp series,

$$\tilde{\nu}_{s,m+1} - \tilde{\nu}_{sm} = F_s(m + 1) - F_s(m) = f(m + \delta). \quad (126f)$$

But, replacing m by $m + \delta$ in (126e), we have

$$f(m + \delta) = F_d(m + 1 + \delta) - F_d(m + \delta).$$

Hence, from (126f),

$$F_s(m + 1) - F_s(m) = F_d(m + 1 + \delta) - F_d(m + \delta),$$

or

$$F_s(m + 1) - F_d(m + 1 + \delta) = F_s(m) - F_d(m + \delta).$$

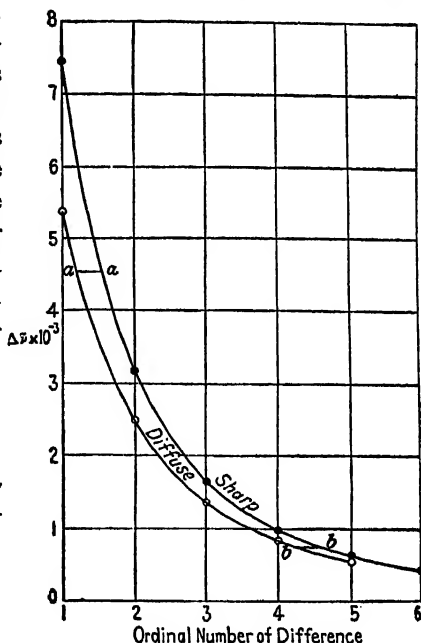


FIG. 61.—Wave-number differences between successive lines in the sharp and the diffuse series of sodium.

Replacing m by $m + 1$ in this equation, we find also that

$$F_s(m+2) - F_d(m+2+\delta) = F_s(m+1) - F_d(m+1+\delta);$$

whence, by the last equation,

$$F_s(m+2) - F_d(m+2+\delta) = F_s(m) - F_d(m+\delta),$$

and, continuing,

$$F_s(m+n) - F_d(m+n+\delta) = F_s(m) - F_d(m+\delta)$$

for any integral n . Hence, replacing $m+n$ by m simply, we can write, for any m ,

$$F_s(m) - F_d(m+\delta) = C,$$

where C is independent of m , or

$$F_s(m) = C + F_d(m+\delta). \quad (126g)$$

Thus, if we know $F_d(m)$ and δ , we can at once write down $F_s(m)$.

The form of F_d could only be discovered by a judicious guess. Rydberg found that many observed series could be represented closely by an equation of the form

$$\bar{\nu}_m = \bar{\nu}_\infty - \frac{R}{(m+\mu)^2}, \quad (127)$$

μ and $\bar{\nu}_\infty$ being constants which vary from one series to another. Obviously, by properly choosing the ordinal number m of the lines, μ can always be made less than 1 in such a formula. The constant $\bar{\nu}_\infty$ represents the high-frequency limit to which the lines in the series ultimately converge. The Balmer formula is a special case of this more general "Rydberg formula," as it is called; for Eq. (125) can be written

$$\bar{\nu} = \frac{c}{\lambda} = \frac{c}{b} - \frac{4c}{bm^2} = \nu_\infty - \frac{R}{m^2},$$

with $\nu_\infty = c/b$, $R = 4c/b$, $m = 3, 4, 5, \dots$ This is of the type of (127) with $\mu = 0$.

The constant R in Eq. (127) was found to be the same for a large group of series for each substance, and very nearly the same for all substances. The slight variation of R from one atom to another is now known to be connected with differences in the atomic weight, the effect of which can be calculated theoretically (cf. Sec. 103). It is thus possible to calculate what the value of R would be for an atom of infinite weight. This latter value of R is denoted by R_∞ ; the value for a particular kind of atom may then be indicated by a subscript.

According to Birge¹ some probable modern values of R , obtained in this manner from spectroscopic observations as represented by Balmer's formula with a slight correction [Sec. 149(b)], are:

For hydrogen $R_H = 109,677.76 \text{ cm.}^{-1}$

For helium $R_{He} = 109,722.40 \text{ cm.}^{-1}$

For infinite atomic weight $R_\infty = 109,737.42 \text{ cm.}^{-1}$

The value of R for the particular atom under discussion is known as the *Rydberg constant* for that atom.

97. Further Relationships between Series. Spectral Terms.—In addition to the qualitative similarities between spectral series already described, Rydberg discovered quantitative relationships between them. For one thing, he noted that the *sharp and the diffuse series appeared to have a common convergence limit*. Next, he observed that this common limit was equal to a *term* in the formula for the *principal series*, viz., to the term $R/(m + \mu)^2$ with m set equal to 1; and that a similar equality held between the *limit* of the *principal series* itself and the variable *term*, for $m = 1$, in the formula for the *sharp series*.

Because of these remarkable relations, the formulas for the three series in question could be written in the following form:

$$\bar{\nu}_p = \frac{R}{(1 + S)^2} - \frac{R}{(m + P)^2} \quad m = 1, 2, \dots \quad (128a)$$

$$\bar{\nu}_s = \frac{R}{(1 + P)^2} - \frac{R}{(m + S)^2}, \quad m = 2, 3, \dots \quad (128b)$$

$$\bar{\nu}_d = \frac{R}{(1 + P)^2} - \frac{R}{(m + D)^2}, \quad m = 2, 3, \dots \quad (128c)$$

Here we have written P, S, D for the values of μ that occur in the variable terms of the formulas for the principal, sharp, and diffuse series, respectively; and we have indicated that, at least for the alkali metals, m starts from 1 in the principal series but from 2 in the others.

Two additional relations involving wave numbers of lines are at once apparent from the formulas. If we set $m = 1$ in the formula for $\bar{\nu}_s$, we obtain the same number as the value of $\bar{\nu}_p$ for $m = 1$, but with reversed sign! Furthermore, the difference between the limit of the principal series and the common limit of the other two is just the wave number of the first line of the principal series (the "Rydberg-Schuster" law, specifically enunciated by Rydberg in 1896 and independently in the same year by Schuster).

The structure of these formulas suggested to Rydberg the possibility that the first term on the right might also vary, in the same

¹ BIRGE, *Rev. Modern Physics*, vol. 1, p. 60 (1929).

manner as does the second, thus giving rise to additional series of lines; for example, we might expect to find a series represented by the formula

$$\bar{\nu} = \frac{R}{(2+S)^2} - \frac{R}{(m+P)^2}, \quad m = 3, 4, \dots$$

Lines or series of this sort were actually discovered later by Ritz. Such lines are called *intercombination lines* or *series*, and the possibility of their occurrence is known as the *Ritz combination principle*. Many examples of them are now known.

The most significant features about atomic spectra in general thus seem to be the following:

1. *The wave number of each line is conveniently represented as the difference between two numbers. These numbers have come to be called terms.*

2. *The terms group themselves naturally into ordered sequences, the terms of each sequence converging toward zero.*

3. *The terms can be combined in various ways to give the wave numbers of spectral lines.*

4. *A series of lines, all having similar character, results from the combination of all terms of one sequence in succession with a fixed term of another sequence. Series formed in this manner have wave numbers which, when arranged in order of increasing magnitude, converge to an upper limit.*

In the further development of spectroscopy, the spectral terms have tended increasingly to become the primary object of study. In the analysis of a new spectrum, one of the first steps is to represent the wave numbers of all of the observed lines as differences between terms, using as few terms as possible. Rydberg's approximate formulas for the spectral frequencies result from the use of an approximate formula for the terms of the type

$$\frac{R}{(m + \mu)^2}, \quad (129)$$

μ being a constant.

The simple picture presented here, however, requires considerable extension in order to be adequate even regarding the spectra emitted by single atoms. For one thing, in writing Rydberg formulas for spectral "lines," we have ignored the fine structure of the lines, by means of which series were first picked out; *singlet* series, in which each line is single, are known in many elements, but more commonly the lines form *doublets*, *triplets*, or groups of even more components. In such cases, a separate Rydberg formula must be written for each

component line; in the spectra of the alkali metals, for example, doublets occur, so that six formulas instead of three are required for a complete representation of the chief series. Furthermore, the Rydberg formula itself is only a first approximation. The whole subject can be much better understood when the discussion can be centered about the atomic mechanism by which spectral lines are emitted. Accordingly, further discussion of details will be postponed to a later chapter.

The remarkable properties possessed by spectral series pointed to the existence within atoms of a comparatively simple and universal mechanism by which spectra are emitted. In terms of classical ideas, however, it was very difficult to imagine a mechanism which could emit spectra having the observed features. It was natural to assume that the higher members of a series were of the nature of overtones. Among acoustic vibrations, many cases are met with in which the frequencies of the overtones are not integral multiples of the fundamental frequency; examples are the vibrations of bells and of the common tuning fork. But no cases are known in which the frequencies of the overtones converge to an upper limit. Furthermore, the Ritz combination principle is without analogy in the classical theory of vibrations.

The key to the origin of spectral lines was not discovered until certain other lines of evidence had led to the adoption of radically new conceptions concerning the structure of atoms.

98. Early Views on Atomic Structure.—Speculations as to the structure of the atom date from the early years of the nineteenth century. In 1815, Prout proposed a hypothesis asserting that *all elements are made up of the atoms of hydrogen as a primordial substance*. The hypothesis was based on the fact that the atomic weights of a large number of the elements are very nearly simple multiples of that of hydrogen, and Prout had no further data to support his views. Accordingly, when more accurate determinations showed that, in general, atomic weights were *not* exact integral multiples of the atomic weight of hydrogen and that there were such notable exceptions as chlorine, with an atomic weight of 35.5, Prout's hypothesis was abandoned—to be revived again decades later in a new form and, of course, on the basis of newly discovered *experimental* evidence (see Chap. XI).

With the discovery of the electron by Sir J. J. Thomson, in 1897, theories of atomic structure began to assume a more definite form, since it became obvious then that the atom must be made up of numerically equal quantities of negative and of positive charges.

Two questions of importance then arose: (1) How many electrons are there in atoms? and (2) How are the electrons and the positive charges in the atom arranged? Two independent lines of evidence gave an answer to the first of these two questions.

1. On the basis of classical theory, Sir J. J. Thomson showed that, when a beam of X-rays passes through matter, it should be scattered, the scattering coefficient σ being given by

$$\sigma = \frac{8\pi}{3} \frac{e^4}{m^2 c^4} N,$$

where e and m are, respectively, the charge and mass of the electron, c is the velocity of light, and N is the number of electrons per unit volume [Sec. 184(a)]. Knowing N from the experimentally determined value of σ and knowing the number of atoms per unit volume of the scattering material, the number of electrons *per atom* could be computed. From early measurements of σ Thomson¹ computed that the number of electrons per atom was of the order of the atomic weight; but, from later measurements,² Barkla concluded that the number of electrons per atom is, for the lighter atoms at least (except hydrogen), more nearly *one-half the atomic weight* [see Sec. 184(b)]. The carbon atom, for example, with atomic weight 12, was found in this way to have 6 electrons. It was later shown that the scattering coefficient for hydrogen is such as to indicate that it has only 1 electron per atom.

2. A stream of electrons traveling with very high velocity is able to pass through thin sheets of matter but, in so doing, is diffused, the electrons being deflected, more or less, from their initial direction as a result of the electrostatic forces acting on the moving electrons when they pass near the electrons in the atoms of the thin sheet. Comparing the computed value of the diffusion with that observed experimentally, Thomson concluded that the number of electrons per atom should be of the order of the atomic weight.

These conclusions were corroborated in a general way by calculations in regard to the dispersion of light by monatomic gases. The number of electrons in an atom was thus not determined definitely, but its order of magnitude was indicated. Since the normal atom is electrically neutral, the quantity of *positive* electricity per atom was thus roughly determined as well. Now, Thomson had shown that the mass of the electron is of the order of one two-thousandth of the mass of the hydrogen atom. It was obvious, therefore, that the

¹ "Electron Theory of Matter," p. 145.

² *Phil. Mag.*, vol. 21, p. 648 (1911).

mass of the atom is vastly greater than the total mass of the comparatively few electrons which it contains; and it was logical to assume that practically the entire mass of the atom is associated with its positive charge.

The problem as to the *arrangement* of the electrons and of the positive charge in the atom then came to the fore. It seemed obvious, on the basis of classical ideas, that this arrangement must be such as to meet two conditions. (1) The ensemble of positive charge and negative electrons which make up the atom must be stable; the electrons, for example, must be held by (electrostatic?) forces in fixed positions of equilibrium about which they may vibrate, when disturbed, with the definite frequencies required to explain the characteristic line spectra of the elements. (2) Except when so disturbed, the electrons must be at rest, since otherwise they would emit radiation as required by the electromagnetic theory. The much greater mass of the positive charge made it reasonable to assume that it is the electrons, rather than the positive charges, which vibrate in the process of emitting radiation.

99. The Thomson Atom.—To meet the requirements of stability and “in default of exact knowledge of the nature of the way in which positive electricity occurs in the atom,” Thomson considered “a case in which the positive electricity is distributed in the way most amenable to mathematical calculation, *i.e.*, when it occurs as a sphere of uniform density *throughout* which the corpuscles are distributed.”¹ The force on a “corpuscle” or electron in such a sphere is easily found from the ordinary laws of electrostatics. Let the charge on the electron be $-e$, that on the sphere E , and the radius of the sphere a . Then, when the electron is at a distance r from the center of the sphere, it experiences a force eEr/a^3 directed toward the center of the sphere. The position of equilibrium with 1 electron is, therefore, *the center of the sphere*, and the *return force* acting on this electron when displaced should be proportional to r . The vibration should be simple harmonic, and an atom so constituted should be capable of emitting only a single frequency, of the order of 10^{15} vibrations per second (using the values $e = 4.77 \times 10^{-10}$ e.s.u., $a = 10^{-8}$ cm., and the mass of the electron, 9×10^{-28} gram). It is, perhaps, suggestive that the order of magnitude of this frequency corresponds roughly to that observed in spectra. But, of course, no atom is known which emits only a single frequency.

Thomson made extensive investigations of the positions of stable equilibrium for various numbers of electrons inside spheres of positive charge. He pointed out that many different frequencies would be

¹ Quotations from Thomson's “The Corpuscular Theory of Matter,” p. 103.

emitted by a system of many electrons in such a sphere but was unable to show that these frequencies might be such as to form a series converging to an upper limit. Eventually his theory came into conflict with the experiments of Rutherford and his collaborators on the scattering of α particles, which are now to be described. These experiments led to the abandonment of the Thomson model of the atom and the substitution of quite a different one.

100. The Scattering of Alpha Particles by Atoms.—The α -rays from radioactive materials have been shown to be positively charged particles which have (1) a mass exactly equal to that of the helium atom and (2) a positive charge numerically equal to twice the charge on the electron. Since helium is known to be produced by radioactive substances which emit α -rays, the latter are identified with helium atoms which have lost 2 electrons (*cf.* Secs. 208 and 209 in Chap. XI). The initial velocity of the α particles, although depending somewhat on the radioactive material from which they originate, is of the order of 2×10^9 cm. per second. These particles can be studied by means of the flashes of light or scintillations which they produce on striking a zinc sulfide screen, the impact of a *single* particle producing a visible flash. These flashes are readily observed under a low-power microscope.

If a stream of α particles, limited by means of suitable diaphragms to a narrow cylindrical pencil, be allowed to strike a zinc sulfide screen placed at right angles to the path of the particles, the scintillations will occur over a well-defined circular area equal to the cross section of the pencil. If, now, a thin film of matter, such as gold or silver foil, is interposed in the path of the rays, it is found that they pass quite readily through the foil but that the area over which the scintillations occur becomes larger and loses its definite boundary, indicating that some of the particles have been deflected from their original direction. This spreading out of the stream of particles on passing through thin layers of matter, solid or gaseous, is called "scattering."

Qualitatively, it is easy to explain the origin of the forces which cause the deflection of the α particle. The particle itself has a twofold positive charge. The atoms of the scattering material contain charges, both positive and negative. In its passage through the scattering material, the particle experiences electrostatic forces the magnitude and direction of which depend on how near the particle happens to approach to the centers of the atoms past which *or through which* it moves.

If we assume the Thomson model of the atom, with its sphere of positive electrification, inside which are electrons, the path of an α

particle in passing *through* such an atom might be as indicated in Fig. 62, the net result of the passage being to deflect the path of the particle through a small angle θ . The major part of this deflection arises from the electrostatic repulsion on the α particle due to the charge on the positive sphere, which, for the heavier atoms at least, has a mass many times that of the α particle. The electrons within the positive sphere, being capable of motion about their respective positions of equilibrium and possessing a mass which is very small compared with the mass of the α particle, will produce no appreciable deflection of the latter. Rather, the electrons themselves would be pulled from their positions of equilibrium and set vibrating or even expelled from the atom by the passage of the α particle. The total deflection of any given particle in passing through or past a number of such atoms in a thin layer of scattering material will be governed by the laws of probability. Thomson showed¹ that the mean deflection ϕ_m of a particle in passing through a thin plate of thickness t should be

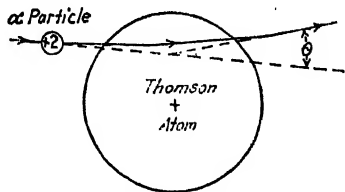


FIG. 62.—Deflection of an α particle by a Thomson atom.

$$\phi_m = \theta \sqrt{N\pi a^2 t},$$

where θ is the average deflection due to a single atom, N is the number of atoms per unit volume, and a is the radius of the positive sphere.

This process of scattering of the α particles as a result of a large number of small deflections produced by the action of a large number of atoms of the scattering material on a single α particle is called *multiple*, or *compound*, scattering. It is readily seen that the structure of the atom assumed by Thomson would not result in a large deflection due to any *single* encounter. According to Rutherford, the number of α particles N_ϕ which, as a result of *multiple* scattering, should be scattered by such an atom through an angle ϕ or greater is given by

$$N_\phi = N_0 e^{-(\phi/\phi_m)^2},$$

where N_0 is the number of particles for $\phi = 0$ and ϕ_m is the average deflection after passing through the scattering material.

Now Geiger had shown² experimentally that the most probable angle of deflection of a pencil of α particles in passing through gold foil 1/2,000 mm. thick is of the order of 1° . It is evident, therefore, from the last equation, that the probability for scattering through

¹ *Cambridge Phil. Soc., Proc.*, vol. 15, p. 465 (1910).

² *Roy. Soc., Proc.*, vol. 83, p. 492 (1910).

large angles becomes vanishingly small; for 30° , for example, it would be of the order of 10^{-13} . Geiger showed that the *observed* scattering obeyed this probability law for *small* angles of scattering but that *the number of particles scattered through large angles was much greater than the theory of multiple scattering predicted*. Indeed, Geiger and Marsden showed¹ that 1 in 8,000 α particles was turned through an angle of *more than* 90° by a thin film of platinum, *i.e.*, was, *in effect*, *diffusely reflected*. This so-called "reflection," however, was shown to be not a surface phenomenon but rather a volume effect, since the number of particles turned through more than 90° increased, up to a certain point, with increasing thickness of the scattering foil. It was also found that the proportion of particles diffusely reflected increased approximately as the $\frac{3}{2}$ power of the atomic weight of the foil.

It was impossible to explain this excess scattering of α particles at large angles on the basis of multiple scattering by a Thomson atom and the laws of probability. There must be something wrong, then, with Thomson's picture of the atom.

101. Rutherford's Nuclear Atom.—(a) *Rutherford's Hypothesis*.—Accordingly, Rutherford, in a classic article² which may be regarded as the starting point of our modern ideas on atomic structure, proposed a new type of atom model capable of giving to an α particle a *large* deflection as a result of a *single* encounter. He assumed *that the positive charge of the atom, instead of being uniformly distributed throughout a sphere of atomic dimensions, is concentrated in a very small region less than 10^{-12} cm. in diameter at the center of the sphere*. This concentrated charge, later called the "nucleus," was assumed to be surrounded by the electrons in some sort of configuration.

The difference in the effect which the two atoms, Thomson's and Rutherford's, have on an α particle passing in their neighborhood is shown in Fig. 63. In (a) is represented diagrammatically the path of an α particle through Thomson's atom, the initial path of the particle being so directed that, if it were not deflected, it would pass through the atom at a distance p from its center. The force which the particle experiences as it penetrates more deeply into the Thomson atom becomes less and less. At point c , the force, although at right angles to the path, is a minimum. The deflection produced, equal to the angle between ab and de , is small. In Rutherford's model [Fig. 63(b)] we have a different state of affairs. Over the path $a'b'$,

¹ *Ibid.*, vol. 82, p. 495 (1909).

² *Phil. Mag.*, vol. 21, p. 669 (1911). Every student of modern physics should read and thoroughly digest this article.

the forces which the particle experiences are the same as for the corresponding path in (a). After passing point b' , however, the forces *continue to increase* according to the inverse square of the distance between the particle and the nucleus, instead of decreasing as was the case in (a). The difference in the forces experienced by the α particle in the two cases becomes very great as the particle approaches the center. It is readily seen that the net result of all this is that the

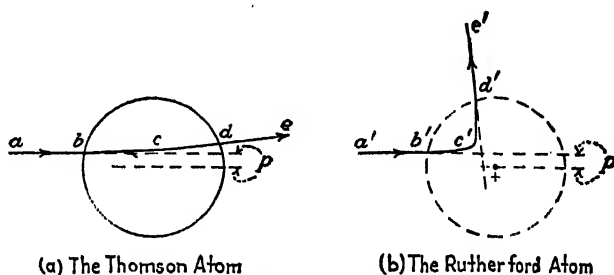


FIG. 63.—Comparison of deflections of an α particle by (a) the Thomson atom and (b) the Rutherford nuclear atom, for similar conditions of incidence.

Rutherford model gives to the particle a much greater deflection than the Thomson model.

Rutherford calculated mathematically the distribution of the particles to be expected, as the result of *single scattering* processes by atoms of the type assumed. The scattering is usually measured by allowing the particles, after passing through the foil F (Fig. 64), to fall on the zinc sulfide screen S placed normal to the initial path of the particles. By means of a low-power microscope, the number of particles striking the screen in the neighborhood of P is observed for various angles of scattering ϕ . Rutherford showed that the number of particles per unit area striking the screen

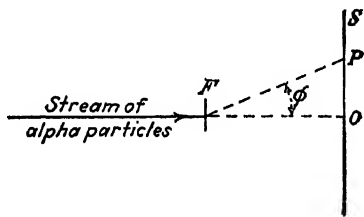


FIG. 64.

should be proportional to (1) $1/\sin^4 \frac{\phi}{2}$; (2) the thickness t of the scattering material, for *small* values of t ; (3) $(Ze)^2$, where e is the charge on the electron and Z is an integer depending on the kind of atom; (4) inversely proportional to the square of the initial kinetic energy of the α particles.

(b) *Experimental Confirmation*.—These predictions were completely verified by the experiments of Geiger and Marsden.¹ Their data are

¹ GEIGER and MARSDEN, *Phil. Mag.*, vol. 25, p. 605 (1913).

shown graphically in Fig. 65, in which the logarithm of the number of scintillations on the screen per minute is plotted as abscissa against the logarithm of $1/\sin^4(\phi/2)$ as ordinate. If these two quantities are proportional to each other, the points for each substance should lie on a straight line inclined at 45° with the axes. The two lines in the figure are drawn at *exactly* 45° , and the observed points are seen to agree well with the predictions. This is the more remarkable since the numbers of scintillations varied in the experiment over a very wide range of values, the left-hand points on the plots representing 22 particles per minute for silver and 33 for gold, whereas the right-hand points represent 105,400 and 132,000, respectively.

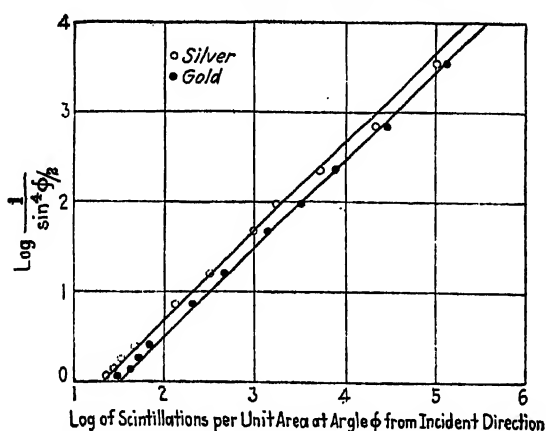


FIG. 65.—The law of the scattering of α particles.

The prediction of Rutherford that the scattering should, for small thicknesses, be proportional to thickness was confirmed by the observations. It will be recalled that, on the theory of multiple scattering by the Thomson atom, the *square root* of the thickness is involved. Geiger and Marsden also showed, in confirmation of the fourth of Rutherford's predictions, that "the amount of scattering by a given foil is approximately proportional to the inverse fourth power of the velocity (inverse square of the energy) of the incident α particles," over a range of velocities such that the number of scattered particles varied as 1:10.

(c) *Atomic Number.*—Geiger and Marsden further concluded, from a study of the variation of scattering with atomic weight and of the fraction of the total number of incident particles scattered through a given angle (1) that the scattering is approximately proportional to the atomic weight of the scatterer over the range of elements from carbon to gold and (2) that "the number of elementary charges

composing the center of the atoms is equal to half the atomic weight." This second conclusion was in agreement with Barkla's experiments on the scattering of X-rays previously mentioned (Sec. 98), which determined the number of *electrons* associated with an atom. According to these results, carbon, nitrogen, and oxygen should have, respectively, 6, 7, and 8 electrons, around a nucleus containing, in each case, an equal amount of charge. Now, these elements are, respectively, the sixth, seventh, and eighth elements in the periodic table. The hypothesis was natural, therefore, that the number of electrons in the atom, *or the number of units of positive charge in its nucleus*, is numerically equal to the ordinal number which the atom occupies in the series of the elements, counting hydrogen as the first. This assumption gives to hydrogen 1 electron, and an equal charge on its nucleus, in agreement with the data on the scattering of X-rays by hydrogen. Helium would then have 2 electrons and a twofold positive charge on the nucleus; and we see, therefore, that the α particles, which have been shown to be helium atoms with a twofold positive charge, are helium nuclei. Lithium, the third element, should have 3 electrons and a threefold positive charge on the nucleus. Neon, the tenth element, should have 10 electrons; and so on. Thus originated the concept of *atomic number*, the importance of which was soon to be emphasized by the pioneer work of Moseley in the X-ray spectra of the elements. The *atomic number*, symbol Z , of an element we may think of variously as (1) the ordinal number of the element in the series of the elements starting with $Z = 1$ for hydrogen, or (2) the positive charge carried by the nucleus of the atom, in terms of the electronic charge e as a unit, or (3) the number of electrons surrounding the nucleus in the normal, neutral atom.

These experiments of Geiger and Marsden so completely confirmed the conclusions which Rutherford had reached by postulating the nuclear type of atom that, in spite of certain weighty objections, the Rutherford atom model was at once universally adopted.

(d) *Some Difficulties.*—The objections to the nuclear type of atom, in which the positive charge occupies a negligibly small volume at the center of the atom and is surrounded by negatively charged electrons in sufficient number to make the atom electrically neutral, are, in large part, concerned with questions of stability. It is obvious that, in the Rutherford atom model, equilibrium cannot be secured, if the electrons are at rest, by the operation of electrostatic forces alone. For, consider [Fig. 66(a)] a nucleus with a double positive charge $+2e$ and with 2 electrons symmetrically placed at a distance r from the nucleus. Assuming the inverse-square law, the attraction of the nucleus for each

electron is $2e^2/r^2$, while the electrons repel each other with a force equal to only one-eighth of this, *viz.*, $e^2/(2r)^2$. The electrons will, therefore, "fall into" the nucleus.

Something is gained by giving to the electron an orbital motion, such as that of the earth round the sun, but not enough. Let a nucleus [Fig. 66(b)] have a charge $+E$ and a mass M , and let a single electron, charge $-e$, mass m very small compared with M , revolve around it in a circular orbit of radius a at such velocity v that

$$\frac{Ee}{r^2} = \frac{mv^2}{r}.$$

We should then have *mechanical equilibrium*. But, according to the fundamental laws of the electromagnetic theory, such a revolving electron, since it is subject to a constant acceleration toward the center of its orbit, should *radiate energy*. This energy can come only from the energy of the system. The system will, therefore, "run down";

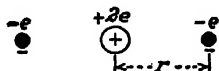


FIG. 66a.

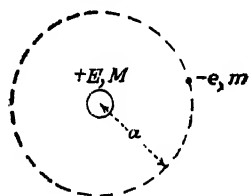


FIG. 66b.

the electron will approach the nucleus by a spiral path and, as can easily be shown, will *give out radiation of constantly increasing frequency*. This, however, is inconsistent with the observed emission of spectral lines of fixed frequency. From the standpoint of stability, the Thomson model was much to be preferred. Yet the experiments of Geiger and Marsden and their interpretation by Rutherford on the basis of the nuclear model of the atom were not to be denied.

It was at this point that Bohr introduced his epoch-making theory of the structure of the atom and of the origin of spectra. His theory constituted an extension of Planck's theory of quanta to Rutherford's nuclear atom, in an attempt, extraordinarily successful, both to remove the difficulties of the nuclear model and to explain the origin of the characteristic spectra of the elements.

102. The Bohr Theory of Atomic Hydrogen.—As we have seen in the last chapter (Sec. 92), the essential features of Planck's original quantum theory were two:

1. An oscillator can exist only in one of a number of discrete quantum states, and to each of these states there corresponds a definite allowed value of its energy.

2. No radiation is emitted while the oscillator remains in one of the quantum states, but it is capable of jumping from one quantum state into another one of lower energy, the energy lost in doing so being emitted in the form of a pulse or quantum of radiation.

Successful applications of the first of these assumptions had already been made in other fields, notably by Einstein and especially by Debye in the specific heat of solids, which will be discussed in a later chapter (Sec. 175, Chap. IX). Nicholson had also attempted to apply the theory to spectra, and with some success, but he was unable to make it yield a series of lines converging to a limit. Bohr discovered how to apply similar ideas to a hydrogen atom of the Rutherford type and succeeded in arriving at a theoretical formula for its spectrum that agrees with observation.¹

Bohr assumed that an electron in the field of a nucleus was not capable of moving along every one of the paths that were possible according to classical theory but was restricted to move along one of a discrete set of allowed paths. While so moving, he assumed that it did not radiate, contrary to the conclusion from classical theory (Sec. 38), so that its energy remained constant; but he assumed that the electron could jump from one allowed path to another one of lower energy and that, when it did so, radiation was emitted containing an amount of energy equal to the difference in the energies corresponding to the two paths.

As to the *frequency* of the emitted radiation, he considered several alternative hypotheses but finally adopted the same assumption that Planck had made for his oscillators. That is, if W_1 and W_2 are the energies of the atom when moving in its initial and final paths, respectively, the *frequency* of the radiation emitted is determined by the condition that

$$\nu = \frac{W_1 - W_2}{h}, \quad (130)$$

where h is Planck's constant (Secs. 92 and 93). This assumption had the additional advantage of agreeing with that made by Einstein in arriving at his highly successful photoelectric equation [Sec. 49, Eq. (53)]. For the latter reason, Eq. (130) came later to be known as the *Einstein frequency condition*.

Concerning the formulation of the condition that determines the allowed paths, Bohr was also in some doubt. While the electron remains in one of its "stationary states," as he called them, he supposed it to revolve in a circular or elliptical orbit about the nucleus, just as the

¹ N. BOHR, *Phil. Mag.*, vol. 26, p. 1 (1913).

earth revolves about the sun, in accordance with the classical laws of mechanics. But what fixes the size and shape of this orbit? Bohr showed that it would suffice to assume a certain relation between the frequency of the emitted radiation and the frequency of revolution of the electron in its orbit. In the end, however, he preferred to postulate that the orbit is a circle, with the nucleus at its center, and of such size that the angular momentum of the electron about the nucleus is equal to an integral multiple of $h/2\pi$.

It was recognized later that the postulate in this form is in harmony with the final interpretation that Planck had found for h in the case of the harmonic oscillator. We saw in Sec. 92 that the state of the

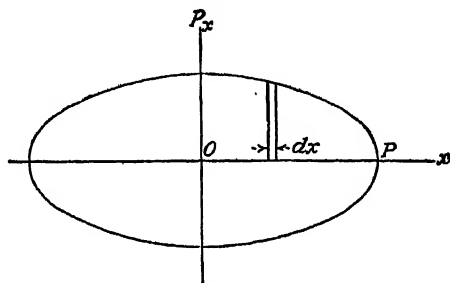


FIG. 67.

oscillator might be represented by a point on a plane, with the coordinate x and the momentum p of the oscillator taken as Cartesian coordinates, and that, as long as the energy of the oscillator remained constant, this point would move along an ellipse (Fig. 67). The condition for a stationary state or quantum state of the oscillator could be said to be that the area enclosed by this ellipse should be an integral multiple of h . For the enclosed area, we can obviously write $\oint p_x dx$, \oint meaning integration over a complete cycle. While x ranges from its minimum to its maximum value, p_x is positive (above Ox in Fig. 67); whereas during the other half-cycle, as x decreases again so that $dx < 0$, p_x is negative, and $p_x dx$ is again positive. Thus a positive value is obtained for the area. Planck's condition for a quantum state of the oscillator can, therefore, be stated as follows:

$$\oint p dx = nh, \quad (131)$$

where n stands for any positive integer.

This rule can at once be modified so as to apply to the revolving electron. If we take, as its coordinate, its angular displacement from a fixed line, ϕ (Fig. 68), the (generalized) momentum p_ϕ corresponding to ϕ is the ordinary angular momentum. Equation (131) thus takes

the form, for the revolving electron,

$$\oint p_{\phi} d\phi = nh.$$

But here p_{ϕ} is constant during a revolution, so that $\oint p_{\phi} d\phi = p_{\phi} \oint d\phi$ and $\oint d\phi = \int_0^{2\pi} d\phi = 2\pi$. Thus, from the last equation, we have, as the condition for a quantum state,

$$p_{\phi} = n \frac{h}{2\pi}, \quad (132)$$

n being any positive integer. This is exactly Bohr's assumption concerning the angular momentum.

The postulates finally adopted by Bohr for the hydrogen atom may thus be summarized as follows:

1. The electron can revolve steadily in a circular orbit, its motion being governed by the ordinary laws of mechanics and electrostatics, provided its angular momentum is an integral multiple of $h/2\pi$. No other orbit is possible (or, at all events, none with different energy).

2. While so revolving, the electron does not radiate. The "radiation reaction" which, according to classical theory, should retard the motion and thereby serve to abstract the radiated energy is also missing.

3. The electron can jump discontinuously from one orbit to another. If it jumps spontaneously into an orbit of lower energy, the energy lost is emitted as radiation whose frequency ν is given by the Einstein frequency condition, Eq. (130) above.

Perhaps the student may be inclined to feel annoyed that this list of assumptions was not written down at once as the basis of the new theory. Our purpose in giving first some further details concerning Bohr's actual line of thought was, in part, to illustrate the fact that confusion usually reigns while important advances in physical knowledge are being made. As a rule, it is only afterward that a neat, logical path can be laid out leading straight to the goal.

Bohr's theory represents a remarkable combination of principles taken over from classical theory with postulates radically at variance with that theory. He solved the old problem of stability merely by *postulating* that the cause of instability, the emission of radiation and the accompanying radiation reaction, did not exist so long as the electron remained in one of its allowed stationary orbits. The electron could thus remain in its stationary state of lowest energy indefinitely,

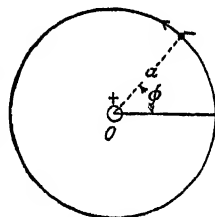


Fig. 68.

without spiraling down into the nucleus, as classical theory would require. But the problem of stability was solved only at the expense of throwing away the only picture we had of the mechanism by which the atom could radiate energy. For Bohr's postulates provide no picture of the sequence of events *during transitions between orbits*.

A hybrid theory of this sort was widely felt to be unsatisfactory; but we shall see in the next section that it was astonishingly successful (in the case of atomic hydrogen). In a later chapter on wave mechanics (Chap. VII), we shall describe the theory that ultimately replaced the Bohr theory. This theory leads to the same set of energy values for the stationary states or quantum states of the hydrogen atom as does the theory of Bohr, but it suggests quite a different picture of the behavior of the electron while in a quantum state. In the new theory, it is only about half true that the electron is in motion in the atom, even when it is in one of its higher quantum states; at least, it cannot be said to follow a definite orbit. Because of the abstractness of the new theory, however, the original simple Bohr picture is commonly felt to retain something more than mere historical interest.

103. Quantum States of One Electron in an Atom.—The allowed values of the energy as given by Bohr's theory for an atom containing a single electron are easily deduced from his basic assumptions. For generality, let the nuclear charge be Ze , Z being the atomic number and, therefore, an integer, and e being the numerical electronic charge, in electrostatic units. Our calculations will then hold not only for the neutral hydrogen atom but also for a singly ionized atom of helium, for doubly ionized lithium, and so on.

In accordance with Bohr's quantum postulate, we assume that the electron revolves in a circle with angular momentum

$$p_{\phi} = ma^2\omega = n \frac{h}{2\pi}, \quad (133)$$

where n is a positive integer, called the "quantum number" of the state in question, m is the electronic mass, a the radius of the orbit, and ω the angular velocity of revolution. To satisfy the laws of mechanics, we must have the necessary centripetal force on the electron supplied by the force of attraction due to the nucleus or

$$ma\omega^2 = \frac{Ze^2}{a^2}. \quad (134)$$

Eliminating ω from these two equations, we find for the radius of the orbit

$$a = \frac{n^2 h^2}{4\pi^2 m e^2 Z}. \quad (135)$$

The energy of the electron will be partly kinetic and partly potential. If we call the energy zero when the electron is at rest at infinity, its potential energy in the presence of the nucleus, according to the usual electrostatic formula (charge times potential), is

$$U = -\frac{Ze^2}{a}.$$

Its kinetic energy is

$$K = \frac{1}{2}mv^2 = \frac{1}{2}ma^2\omega^2 = \frac{1}{2}\frac{Ze^2}{a},$$

by (134). Thus $K = -\frac{1}{2}U$, a relation peculiar to motion under an inverse-square force. The total energy is $W = T + U$ or

$$W_n = -\frac{1}{2}\frac{Ze^2}{a} = -\frac{2\pi^2me^4Z^2}{h^2n^2}, \quad (136)$$

by (135). This is the energy of the atom when the electron is in its n th stationary or quantum state. Here and in Eq. (135), the quantum number n may take on any integral value: $n = 1, 2, 3, \dots$

From Eq. (136), we see that the larger the value of n , the *smaller numerically* but the larger in *algebraic value* is the energy of the system. The lowest value of W_n is that corresponding to the *first* orbit. This is known as the *normal* (quantum) state, or *ground state*, of the atom, since it should be the *most stable state and the one ordinarily occupied by the electron*.

It is interesting to note that according to Eq. (135) the radii of the successive allowed orbits are proportional to n^2 , to $1^2, 2^2, 3^2, 4^2, \dots$ or to $1, 4, 9, 16, \dots$. If in (135) we substitute $Z = 1$ for hydrogen and $h = 6.61 \times 10^{-27}$, $e = 4.803 \times 10^{-10}$, $m = 0.9107 \times 10^{-28}$, and $n = 1$, we find for the radius of the smallest Bohr circle for hydrogen, or the radius of its orbit in the normal atom,

$$a = a_0 = \frac{h^2}{4\pi^2me^2} = 5.27 \times 10^{-9} \text{ cm.} \quad (137)$$

The diameter of the orbit is thus close to 10^{-8} cm., which agrees very well with estimates of the atomic diameter obtained from kinetic theory. This is a first indication that the new theory may be able to explain, among other things, the apparent sizes of molecules. In atoms containing only 1 electron but a larger nuclear charge, the orbits are all smaller in the ratio $1/Z$.

It must be remarked, however, that our calculations contain a slight error. What we have actually developed is a *fixed-nucleus* theory. In reality, if the electron revolves in a circle about the nucleus, as we

have supposed it to do, its path in space will be a circle about the center of mass of the combined system, which is slightly displaced from the nucleus (Fig. 69). At the same time the nucleus revolves about

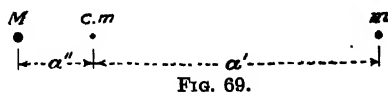


FIG. 69.

the center of mass in a smaller circle. If a is the distance of the electron from the nucleus, and a' , a'' the respective distances of electron

and nucleus from the center of mass, then

$$a' = \frac{M}{m+M}a, \quad a'' = \frac{m}{m+M}a,$$

where M is the mass of the nucleus.

The total angular momentum about the center of mass is then the sum of that due to the electron and that due to the nuclear motion or

$$ma'^2\omega + Ma''^2\omega = \left[m \frac{M^2}{(m+M)^2} + M \frac{m^2}{(m+M)^2} \right] a^2\omega = \frac{mM}{m+M}a^2\omega.$$

Let us write

$$m' = \frac{mM}{m+M}; \quad (138)$$

m' so defined is called the *reduced mass* of the electron. Then we can write for the total angular momentum $m'a^2\omega$, and, if we assume that Bohr's postulate has reference to the total angular momentum of the atom, we shall have, in place of (133),

$$m'a^2\omega = n \frac{h}{2\pi}. \quad (133')$$

Furthermore, in the left-hand member of the dynamical equation, (134), a , representing the radius of the electronic orbit in space, is now to be replaced by a' . The same result is obtained, however, if, retaining a , we replace m by m' as defined by Eq. (138). In the *right-hand* member of (134), however, a represents the distance between the electron and the nucleus and must be left unchanged. Thus in place of (134) we have

$$m'a\omega^2 = \frac{Ze^2}{a^2}. \quad (134')$$

The expression for the potential energy U is unaltered. The kinetic energy of the electron, however, will now be $\frac{1}{2}m'a^2\omega^2$, to which is to be added that of the nucleus, or $\frac{1}{2}Ma''^2\omega^2$; the total kinetic energy of the atom is thus

$$\begin{aligned} \left(\frac{1}{2} m a'^2 + \frac{1}{2} M a''^2 \right) \omega^2 &= \frac{1}{2} \left[m \frac{M^2}{(m+M)^2} + M \frac{m^2}{(m+M)^2} \right] a^2 \omega^2 \\ &= \frac{1}{2} \frac{mM}{m+M} a^2 \omega^2 = \frac{1}{2} m' a^2 \omega^2 = \frac{1}{2} \frac{Ze^2}{a}, \end{aligned}$$

by (134'). The total energy is thus

$$W_n = - \frac{1}{2} \frac{Ze^2}{a}$$

as before.

It is easily seen now that the effect of these changes is simply to replace m by m' as given by Eq. (138) in all of the equations previously written, including Eq. (135) giving the radius of the orbit and Eq. (136) giving the energy, so that these latter equations become

$$a = \frac{n^2 h^2}{4\pi^2 m' e^2 Z}, \quad W_n = - \frac{2\pi^2 m' e^4 Z^2}{h^2 n^2}. \quad (135', 136')$$

We note that the orbital radii and the allowed energies are caused by the nuclear motion to vary slightly with M , or with the atomic weight.

In addition to the *discrete set* of allowed orbits just described, in any one of which the electron is capable of moving with negative total energy, Bohr assumed that it could also move with *any value whatever of positive energy*. In such a case the orbit would be a hyperbola, as some of the orbits of comets have been supposed to be. Thus we have, in all, a *discrete set* of allowed *negative* energies, or energy levels, converging to the value zero and, also, from zero up, a *continuum* of allowed *positive* energies.

The zero value of the energy has been so chosen here that it corresponds to the electron at rest at infinity. In this state the atom can be regarded as just barely ionized. Accordingly, *the numerical value of the (negative) energy of the normal state also represents the least energy required to ionize the atom by removal of the electron.*

A final remark should be added to forestall misunderstanding. The modern wave-mechanical theory assigns definite values of angular momentum to the various quantum states of atoms, as did Bohr's theory, but these values are not the same as those assigned by the Bohr theory [cf. Sec. 124(b)]. The *energy levels*, however, turn out to be just the same, for a 1-electron atom, as those given by Bohr's theory, so long as all relativistic effects (including spin) can be neglected (Sec. 149).

104. Spectrum of a One-electron Atom.—According to Bohr's postulates, a hydrogen atom radiates when the electron jumps from one

quantum state into another state of lower energy. The difference in the two energies is simultaneously emitted as a single quantum of radiant energy the frequency of which is given by the Einstein relation or

$$\nu = \frac{W_1 - W_2}{h}. \quad (139)$$

Inserting here for W_1 and W_2 the values given by Eq. (136) for states with quantum numbers n_1 and n_2 , we have, therefore, for the frequency of the line emitted when the atom jumps from state n_1 to state n_2 ,

$$\nu = \frac{2\pi^2 me^4 Z^2}{h^3} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right), \quad (140a)$$

or, in terms of wave numbers,

$$\bar{\nu} = RZ^2 \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right), \quad R = \frac{2\pi^2 me^4}{ch^3}. \quad (140b,c)$$

These expressions are positive, since necessarily $n_1 > n_2$.

For hydrogen ($Z = 1$), the latter formula becomes

$$\bar{\nu} = R \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right). \quad (141)$$

If we set $n_2 = 2$ and let n_1 take on any integral value from 3 up, this last formula is exactly of the type of Balmer's formula! It permitted at once a crucial quantitative test of the new theory. When Bohr inserted in it the best values of his day, $e = 4.7 \times 10^{-10}$, $e/m = 5.31 \times 10^{17}$ with e in electrostatic units, and $h = 6.5 \times 10^{-27}$ (as found from Planck's formula for black-body radiation), he found for the coefficient in (140a)

$$\frac{2\pi^2 me^4}{h^3} = 3.1 \times 10^{15}.$$

This, then, should be the value of the Rydberg constant R , when expressed in terms of frequency rather than wave number. As calculated from the frequencies of the lines of the Balmer series, its value was

$$3.290 \times 10^{15}.$$

Bohr considered that "the agreement between the theoretical and observed values is inside the uncertainty due to experimental errors in the constants entering in the expression for the theoretical value."

If we employ the best modern values, say $e = 4.803 \times 10^{-10}$, $e/m = 1.759 \times 2.9980 \times 10^{17}$ with e in electrostatic units, and

$h = 6.61 \times 10^{-27}$, we obtain, in reciprocal centimeters,

$$\frac{2\pi^2 me^4}{ch^3} = 1.105 \times 10^5 \text{ cm.}^{-1},$$

as against the best spectroscopic value¹ (for an atom of infinite mass):

$$R_\infty = 109,737.42 \text{ cm.}^{-1}.$$

The difference is about 0.7 percent. This would seem to be too large a difference to be due to errors in the experimental values of e , e/m , c , and h . The cause of this discrepancy is at present unknown. The wave-mechanical form of quantum theory leads to the same theoretical formula for R_∞ as the Bohr theory [Eq. (140c)]; and wave mechanics has been so widely successful that its general correctness can hardly be doubted. Perhaps some small further correction, as yet unknown, should be made in the theory. On the other hand, it is a general fact that, as the refinement of experimental methods has increased, the observed values of physical constants have shown in the past a tendency to shift by larger amounts than would have been anticipated in view of the precision of the methods employed, presumably owing to unsuspected systematic errors. The possibility cannot be excluded that errors still exist in the measured values of e and h sufficient to account for the discrepancy in the Rydberg constant.

The maximum possible value of $\tilde{\nu}$ in any given series is obtained from Eq. (140b) if we set $n_1 = \infty$. This gives the theoretical series limit,

$$\tilde{\nu}_\infty = \frac{RZ^2}{n_2^2}. \quad (142)$$

For the corresponding transition, we have $W_1 = 0$, the electron starting from a state of rest at infinity.

The range of possible transitions is not thereby exhausted, however. It can also be imagined that an electron moving along a hyperbolic orbit past a hydrogen nucleus might drop into one of the closed orbits that are allowed on the Bohr theory, emitting all of its excess energy in a single quantum. In this case, W_1 would have some *positive* value W and the frequency of the emitted photon would be, obviously,

$$\tilde{\nu} = \frac{W - W_2}{ch} = \tilde{\nu}_\infty + \frac{W}{ch}.$$

Since W can vary continuously from 0 upward, such jumps would

¹ BIRGE, *Phys. Rev.*, vol. 54, p. 972 (1938).

give rise to a continuum extending from ν_∞ toward shorter wave lengths. Such a continuum is clearly visible in Fig. 59 (Sec. 95).¹

In spite of the radical nature of the assumptions underlying Bohr's theory, its quantitative successes in explaining many facts quickly secured general acceptance of the theory. In his first paper in 1913, Bohr points out several other respects in which the predictions of his theory were in agreement with experiment. According to the theory, in the formulas for ν or $\bar{\nu}$ the quantum number n_2 , as well as n_1 , might have any value. If we make $n_2 = 1$ and let n_1 vary, we obtain another series of lines, for which the final state is the normal state of the atom; $n_2 = 3$, with $n_1 > n_2$, gives us a third series, and so on. In Bohr's time a series of infrared lines corresponding to $n_2 = 3$, called the "Paschen series," was actually known. The series for $n_2 = 1$, however, lies in the ultraviolet and was not discovered until later; it is called, after its discoverer, the "Lyman series." Other series corresponding to $n_2 = 4$ or 5 have also been observed.

It is important to note clearly, however, which features of the theory are directly effective in determining the frequencies of the spectral lines. These frequencies are quite distinct from the *frequency of orbital revolution* of the electron, a feature of the new theory that is in most striking contrast with classical ideas. In order to see just what the relation is between the emitted frequencies and the orbital frequency, we may solve (133) and (134) in the last section for ω instead of a . We thus find for the frequency of orbital revolution,

$$\nu_{\text{orb}} = \frac{\omega}{2\pi} = \frac{4\pi^2 m e^4 Z^2}{h^3} \frac{1}{n^3}.$$

For comparison, Eq. (140a) may be written thus:

$$\nu = \frac{4\pi^2 m e^4 Z^2}{h^3} \frac{n_1 + n_2}{2n_1^2 n_2^2} (n_1 - n_2).$$

It can easily be seen that, since $n_1 > n_2$,

$$\frac{1}{n_1^3} < \frac{n_1 + n_2}{2n_1^2 n_2^2} < \frac{1}{n_2^3}.$$

Hence, if $n_1 - n_2 = 1$, the frequency ν of the emitted radiation is intermediate between the frequencies of orbital revolution in the initial and final states. Only for very large n , for which the orbital

¹ The continuum actually overlaps the series slightly in this figure. This may be due to some interference of the atoms with each other; in its higher quantum states, according to Eq. (135), the hydrogen atom must be effectively very large, e.g., $400 a_0$ or 2×10^{-6} cm. for $n = 20$.

frequencies in successive orbits become indistinguishable, do the emitted and orbital frequencies tend to coincide. Making in succession $n_1 - n_2 = 1, 2, 3, \dots$, we have an approximation to various harmonic overtones of a fundamental frequency. Thus, quantum jumps in which $\Delta n > 1$ correspond to the overtones in the case of classical vibrations. This is an example of a principle later elaborated by Bohr and known as the "correspondence principle."

In the present section we have discussed principally the spectrum of hydrogen, but similar results follow from Eq. (140a) or Eqs. (140b,c) for any atom containing a single electron. As a further example, the spectrum of ionized helium will be discussed presently (Sec. 106). In all cases each spectral frequency emitted is determined solely by the *energies in two stationary states* together with the *Einstein frequency condition*. The most fundamental feature of the theory for spectral emission is, therefore, the *set of allowed energies*.

105. The Spectrum of Atomic Hydrogen. Energy Levels and Spectral Series.—The energies of the lowest eight states of the hydrogen atom, calculated from Eq. (136), are shown in Table II. It is often convenient to express such energies in wave-number units; the difference of two energies then gives at once the wave number of the corresponding spectral line. If W_e is an energy in ergs, its value W_s in wave-number units or cm.^{-1} is

$$W_s = \frac{W_e}{ch} \quad (143)$$

TABLE II. — ENERGY VALUES OF SOME STATES OF THE HYDROGEN ATOM

State	Energy		
	Wave-number units	Ergs	Electron volts
$n = 1$	— 109,678	— 217.3×10^{-13}	— 13.58
2	— 27,420	— 54.3	— 3.394
3	— 12,186	— 24.2	— 1.508
4	— 6,855	— 13.58	— .849
5	— 4,387	— 8.69	— .543
6	— 3,047	— 6.04	— .377
7	— 2,238	— 4.44	— .277
8	— 1,714	— 3.40	— .212

For other purposes energies expressed in *electron-volts* are more convenient. An *electron-volt* (abbreviation, ev) is the work done on an electron, or its gain in energy, when it passes through a potential rise of 1 volt. Hence, a volt being $\frac{1}{300}$ electrostatic unit,

$$1 \text{ electron-volt} = \frac{e}{300} = 1.601 \times 10^{-12} \text{ erg.} \quad (144)$$

if $e = 4.803 \times 10^{-10}$ e.s.u. In Table II, the energies are given in all three units, ergs, cm.^{-1} , and electron-volts.

In spectroscopic work, however, it is more usual not to employ the energy itself, which is negative, but its numerical value. These numerical values of the atomic energies are called *term values* or *terms*. The wave number of a spectral line is then obtained by subtracting the term value for the *initial* state from that for the *final* state.

TABLE III.—SOME TERM VALUES AND LINES FOR THE HYDROGEN ATOM

λ , angstroms	$\tilde{\nu}$, cm.^{-1}	Terms, cm.^{-1}	Quantum number, n
Lyman series		109,678	1
1,216.0	82,258	27,420	2
1,025.8	97,491	12,186	3
972.5	102,823	6,855	4
949.5	105,291	4,387	5
Balmer series		27,420	2
6,562.8	15,233	12,186	3
4,861.3	20,565	6,855	4
4,340.5	23,032	4,387	5
4,101.7	24,373	3,047	6
Paschen series		12,186	3
18,756	5,331	6,855	4
12,821	7,799	4,387	5
10,939	9,139	3,047	6
10,052	9,948	2,238	7
Brackett series		6,855	4
4.05 μ	2,468	4,387	5
2.63	3,808	3,047	6
2.16	4,617	2,238	7
1.94	5,141	1,714	8

In Table III are shown the first four lines of each of the four known series in the spectrum of atomic hydrogen, and the associated term values. Column 4 gives the quantum number for each term. The first term value given in column 3 is in each case the convergence wave

number or limit for that series; the wave number $\bar{\nu}$ of each line is obtained by subtracting from the convergence limit the term opposite the wave number of the line. The wave length λ is calculated as $\bar{\nu}^{-1} \times 10^8$. Many repetitions of term values necessarily occur in such a table.

The relations between the energies and the series of lines are much better seen from an *energy-level diagram*. In Fig. 70(a) is shown such a diagram, necessarily incomplete, for atomic hydrogen. Each horizontal line represents an energy level, higher energies being plotted above. The line at $n = 1$ represents the normal state of the atom; the line at $n = \infty$ represents the electron at rest at infinity, the atom being just ionized; and above this is the continuum of positive energies for the free electron. Wave numbers are shown on a scale reading downward. Energies measured from the normal state as zero are also shown, expressed in electron-volts. A few of the transitions which give rise to lines are shown in an obvious manner by arrows. The diagram brings out clearly the fact that each series ends on a particular energy level.

106. Ionized Helium.—A helium atom which has lost both electrons is a bare nucleus and cannot radiate energy. One which has lost only a single electron, however, resembles a hydrogen atom, except that $Z = 2$ and the nucleus is nearly four times as heavy. The spectrum emitted by such atoms is known as the *spark spectrum* of helium, because it is emitted much more strongly when the helium is excited by a spark than when it is excited by an arc. The *arc spectrum* of helium, emitted by the neutral atom, will be considered later (Sec. 150).

The general theory derived in Sec. 104 should apply at once to ionized helium. Putting $Z = 2$ in Eq. (140b) we should have for the frequencies of the spectral lines

$$\bar{\nu} = 4R \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right). \quad (145)$$

Thus *ionized helium should emit the same spectrum as hydrogen except that all frequencies are four times as great, or all wave lengths a quarter as great.*

This conclusion from the theory agrees with observation except for a slight numerical discrepancy which becomes significant when measurements of precision are considered. As explained in Sec. 103, m in Eq. (136), or in formula (140c) for the Rydberg constant, must, strictly speaking, stand for the reduced mass m' of the electron as given by Eq. (138). Hence, the Rydberg constant for a 1-electron

atom the nucleus of which has a mass M is, from (138) and (140c),

$$R = \frac{M}{m + M} R_{\infty}, \quad R_{\infty} = \frac{2\pi^2 m e^4}{ch^3}. \quad (146a, b)$$

In these latter formulas, m stands for the ordinary electronic mass, 0.911×10^{-27} gram.

One use that can be made of these relations is to deduce the ratio of the mass of the electron to the mass of a hydrogen or helium atom from values of R calculated from spectral data. If subscripts H, He indicate values of R referring to hydrogen and helium, respectively, we have, from (146a),

$$\frac{R_{\text{He}}}{R_{\text{H}}} = \frac{M_{\text{He}}}{m + M_{\text{He}}} \frac{m + M_{\text{H}}}{M_{\text{H}}} = \frac{1 + (m/M_{\text{H}})}{1 + (m/M_{\text{He}})},$$

whence

$$\left(1 - \frac{M_{\text{H}}}{M_{\text{He}}} \frac{R_{\text{He}}}{R_{\text{H}}}\right) \frac{m}{M_{\text{H}}} = \frac{R_{\text{He}} - R_{\text{H}}}{R_{\text{H}}}.$$

If we insert here, from Sec. 96, $R_{\text{H}} = 109,677.76$, $R_{\text{He}} = 109,722.40$, and $M_{\text{H}}/M_{\text{He}} = 1.0078/4.002$, we find

$$\frac{m}{M_{\text{H}}} = \frac{1}{1,838}.$$

That a value of m/M_{H} so found really possesses some precision, although it depends upon the difference of two numbers, R_{H} and R_{He} , which are almost equal, is due, of course, to the extraordinary precision of the spectroscopic values of R . We can then also proceed further and obtain a value of e/m for the electron. This is one of several spectroscopic methods for finding e/m . For

$$\frac{e}{m} = \frac{e}{M_{\text{H}}} \frac{M_{\text{H}}}{m},$$

and e/M_{H} equals the charge necessary to precipitate a gram-atom in electrolysis divided by the atomic weight of hydrogen, or 9,648.9 electromagnetic units divided by 1.0078. Thus,

$$\frac{e}{m} = \frac{9,648.9}{1.0078} \times \frac{1}{1,838} = 1.760 \times 10^7 \text{ e.m.u. gram}^{-1}.$$

This value of e/m agrees very closely with the results of recent measurements of the same quantity for *free* electrons. Birge cites the average of four such measurements, made either by measuring the velocity imparted to electrons by a known potential drop or by the magnetic-

deflection method, as 1.75955.¹ The agreement of the values of e/m obtained by these two different methods constitutes a valuable check on the correctness of the assumptions underlying atomic theory. Furthermore, five other spectroscopic methods of evaluating e/m , listed by Birge, give values in good agreement with these two. The weighted average of all ten values is

$$\frac{e}{m} = 1.7591 \times 10^7.$$

From the value just found for m/M_H we can calculate R_∞ from the observed value of R_H by means of the formula, a special case of Eq. (146a),

$$R_\infty = \left(1 + \frac{m}{M_H}\right) R_H.$$

It is in this or an equivalent way that "experimental" values of R_∞ are obtained.

In a similar way, *doubly ionized lithium* is found to emit the hydrogen spectrum with all frequencies multiplied (almost exactly) by 9; *trebly ionized beryllium* emits them increased in the ratio 16; and so on. The first line of the Lyman series for quadruply ionized boron ($Z = 5$) has been found by Edlén at 48.585 Å, with a frequency 25.04 times that of the first Lyman line of hydrogen.

107. Energy Levels and Series Relationships for Sodium.—Energy-level diagrams analogous to that for hydrogen can be constructed for all atoms. Such a diagram exhibits very clearly a simple explanation, in terms of the energy levels, for those striking properties of spectral series and of their interrelations which were described above (Secs. 96 and 97). As an illustration we may return here to a further brief discussion of the spectrum of neutral sodium.

A partial energy-level diagram for sodium is shown in Fig. 70(b); it contains levels only for those quantum states that occur as the valence electron moves out toward infinity (for a more complete treatment of the sodium spectrum see Secs. 142, 143, 147). The levels are labeled on the diagram in modern spectroscopic notation, but this is immaterial for our present purpose. Most of the levels are really double, two levels lying very close together, but only one such pair is shown as such on the diagram (and on a much exaggerated scale), the others being represented by a single line each. (A similar but more minute fine-structure of the hydrogen levels has been ignored; see

¹ BIRGE, *Phys. Rev.*, vol. 54, p. 972 (1938).

Sec. 149 below.) The levels labeled S , however, including the normal state, are strictly single.

A brief inspection of the level diagrams in Fig. 70 reveals the following explanation of the properties of the spectral series. That each series must converge to a finite limit arises from the simple fact that the largest energy difference possible between levels is limited

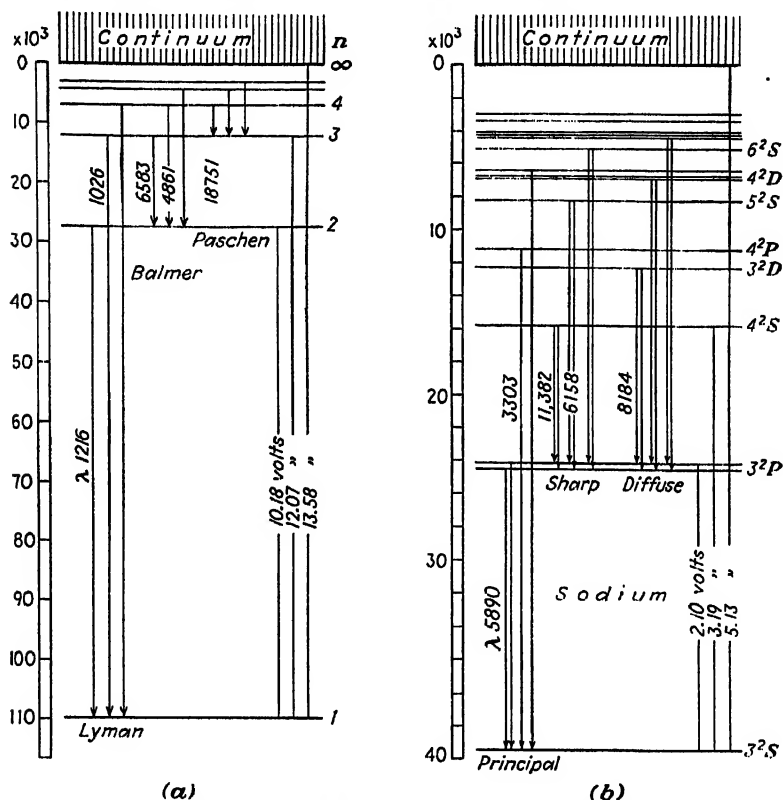


FIG. 70.—The lower energy levels of (a) hydrogen and (b) sodium atoms. Values at the left are in cm^{-1} ; a few differences are shown at the right in electron-volts. The transitions giving rise to the first three lines of the chief series are shown by arrows.

to the ionization energy of the atom (when it is in the initial state for the series in question). The *principal* series of lines for sodium, including as its first member the ordinary D lines, arises from transitions ending on the normal state, which, as just stated, is single. The doublet character of the lines of this series is due, therefore, to the doublet nature of the upper level; and, since the spacing of these levels decreases rapidly with increasing energy, the spectral doublets close up as their ordinal number increases. The *sharp* and the *diffuse* series

both end on the doublet level next above the normal state; hence, they have a common convergence limit. The lines of these series, as observed, usually appear to be doublets of constant frequency difference, this difference being that between the sublevels of the final state; hence, Hartley's law commonly appears to be obeyed by both series. (A further slight splitting of the *diffuse* lines will be discussed later in Sec. 147.) Finally, the basis for the Rydberg-Schuster law is simple and obvious: the difference between the common convergence limit of the sharp and the diffuse series and the limit of the principal series is just the difference between the lowest two energy levels; and this is also equal to the wave number of the first line of the principal series.

The convergence limit of the principal series, corresponding to a jump of the electron from infinity into the normal state, obviously corresponds to an energy that is (1) numerically equal to that of the normal state, when this is measured downward from zero at infinity, and (2) equal to the minimum energy that must be given to the atom to remove an electron from it, *i.e.*, to the *ionization energy* of the neutral atom. The ionization energy is usually expressed in electron-volts and called the *ionization potential*, representing the potential through which an electron must drop in order to acquire enough energy to ionize the atom by impact. It is 5.13 volts for sodium, 13.58 volts for a hydrogen atom.

The discussion of spectra in terms of energy levels is thus much simpler than the direct discussion in terms of the lines themselves. Whatever may happen in the future to our picture of the behavior of the electrons in an atom, energy levels are undoubtedly here to stay.

108. Excitation and Ionization of Atoms by Electrons.—In the last few sections, we have ascribed the emission of radiation to transitions made by atoms from one stationary state to another of lower energy, without considering how the atoms get into the higher states initially. Since the energy of the atom decreases with each emission of a quantum of radiation, the final result of emission processes can only be that the atom arrives in its normal state, or state of lowest energy, and then remains there without radiating farther. Thus, in order to be able to emit radiation, the atom must first be transferred by some means from its normal state into a state of higher energy.

When an atom is in such a state of higher energy, it is said to be *excited*, and the process of transferring or "raising" it from its normal state into an excited state is referred to as *excitation*. The excited atom may be thrown into any one of its infinite series of discrete quantum states. It may also happen, however, that the process of excitation is so intense that the electron is completely removed from

the nucleus. The atom is then left with a net positive charge and, if free, will tend to move in an electric field, like an ion in an electrolytic solution. Such an atom which has lost 1 electron (or more) is said to be in an *ionized* state, and the process of raising the atom from the normal state to the ionized state is called *ionization*. In the process either of excitation or of ionization, additional energy is given to the atom.

A full discussion of the subject of the excitation and ionization of atoms, and of the far-reaching conclusions drawn from experiments in this field, is beyond the scope of this book. The reader is referred to more exhaustive or special treatises. We may, however, make brief mention of some of the ways by which experiment shows that atoms

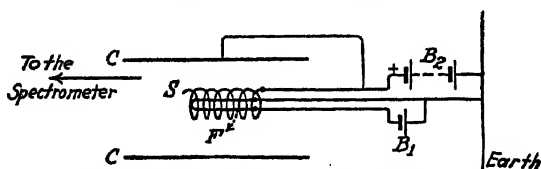


FIG. 71.—The apparatus of Foote, Meggers, and Mohler for studying excitation potentials.

may become excited or ionized and of the interpretation of some of these experiments.

The atoms of a gas or vapor may become excited or ionized by bombarding them with electrons. If the energy given to an atom by such a collision is sufficient to expel an electron from its normal orbit in the atom to infinity, the atom becomes ionized; otherwise, it becomes simply excited.

As one illustration of the mode of procedure followed in studying phenomena of this kind may be mentioned an experiment by Foote, Meggers, and Mohler.¹ Their apparatus is shown diagrammatically in Fig. 71. A filament *F* of tungsten or lime-coated platinum is heated by a battery *B*₁ to such temperature that it emits electrons. Around the filament is a spiral grid *S*, which, by means of a battery *B*₂ or potentiometric source of *pd*, is maintained at any desired *positive* potential with respect to the filament. Around *S* and electrically connected thereto is a metal cylinder *CC*; *S* and *CC* are thus at the same potential. Inside *CC* is the gas or vapor under study, maintained at a suitable pressure. Electrons, emitted by the filament, are accelerated toward the grid *S*, which is so close to *F* that with proper regulation of the gas pressure comparatively few atoms of the gas are

¹ FOOTE, MEGGERS, and MOHLER, *Astrophys. J.*, vol. 55, p. 145 (1922). Cf. also FOOTE and MOHLER, "The Origin of Spectra."

struck by the electrons in their passage from filament to grid. After leaving the grid, the electrons move in the force-free space between *S* and *CC*, in which space, because of the greater distance *S* to *CC*, the electrons collide with the gas molecules causing the excitation of the latter, provided the energy of the electrons is sufficiently great.

With sodium vapor inside the cylinder, no luminosity is observed in the vapor until the potential difference between the grid and the filament reaches 2.09 volts. For voltages slightly above this value, the spectrograph shows that the sodium vapor emits the well-known *D* lines, and those only. The mean wave length of these lines is 5,893 angstroms, corresponding to a frequency ν of 0.509×10^{15} sec.⁻¹; the energy quantum $h\nu$ is

$$6.61 \times 10^{-27} \times 0.509 \times 10^{15} = 3.36 \times 10^{-12} \text{ erg.}$$

According to the quantum picture of the origin of spectra, there must have been a *change of state* of the sodium atom in *emitting* this line corresponding to an *energy drop* of 3.36×10^{-12} erg—whatever may have been the absolute values of the energies in the initial and the final state. Now, the electrons, which, after emission from the filament, have dropped through 2.09 volts, have acquired a kinetic energy given by

$$eV = 4.80 \times 10^{-10} \times \frac{2.09}{3 \times 10^9} = 3.34 \times 10^{-12} \text{ erg.}$$

We see, therefore, that the kinetic energy possessed by the electrons as they pass through the space between the grid and the cylinder is equal to the quantum energy of the radiation emitted by the sodium vapor. The presumption is very strong, therefore, that the electrons have, by collision, transferred to the sodium atoms sufficient energy to raise the latter from their *normal* state to an excited state

$$3.36 \times 10^{-12} \text{ erg}$$

above the normal, so that the atoms in returning from that excited state to the normal state emit the doublet *D* line.

In discussing such observations, it is convenient to express all spectral frequencies in equivalent electron energies. If the quantum energy $h\nu$ of a line is equal to the energy acquired by an electron in falling through *V* volts or *V*/300 electrostatic units of potential, we have

$$\frac{Ve}{300} = h\nu = hc\bar{\nu} = \frac{hc}{\lambda}$$

Inserting $e = 4.803 \times 10^{-10}$, $h = 6.61 \times 10^{-27}$, $c = 2.998 \times 10^{10}$,

and multiplying λ by 10^8 to convert to angstroms, we obtain the convenient conversion formulas

$$\tilde{\nu} = 8,079V, \quad \lambda = \frac{12,378}{V} \text{A.} \quad (147a,b)$$

The quantity V is a measure of the energy of the radiant quantum, or of the energy lost by the atom in emitting it, in terms of the *electron-volt* as a unit.

To return to the experiment of Foote, Meggers, and Mohler, it was found that, as the voltage between filament and grid was raised above 2.09 volts, the D lines continued to appear, but no others appeared until the voltage reached 5.12 volts, beyond which the complete arc spectrum was produced. Beyond 35 volts, the so-called "spark" spectrum was seen. The explanation of these facts cannot be quite the same as for the D lines. For electrons possessing energy equivalent to, say 3.0 volts, should, so far as *energy* is concerned, be able to communicate to sodium atoms sufficient energy to generate lines of wave length 4,130 angstroms or longer. In the spectrum of sodium is found a large number of lines of longer wave length than this, for example practically the entire sharp and diffuse series, the convergence wave length of which is of the order of 4,100 A. Why do not 3-volt electrons excite these lines? Or, for that matter, the wave length of the second line of the sharp series is 6,158 A or 2.01 electron-volts. Why is this line not produced at even lower voltages than the D lines, which correspond to 2.09 volts?

A little study of the energy-level diagram of sodium in Fig. 70(b) above reveals the answers to these questions. A given line can be emitted only if atoms are present in the state which is the *initial* state for that line. To emit the line at 6,158 A, the atom must start from a state with an energy much above the initial state for the D lines, this state being actually the *final* state for the former line; and 2.09-volt electrons cannot possibly put the atom into so high a state. Even 3-volt electrons cannot excite any state above 3^2P , from which the D lines start. On the other hand, since the convergence wave number of the principal series of sodium is 41,450, equivalent to 5.13 volts, an electron with more than 5.13 volts of energy can remove an electron from the atom to infinity, *i.e.*, it can *ionize* the atom. Subsequently, an electron will be picked up by the ionized atom, dropping from one discrete level into another with the emission of radiation until it reaches the lowest level and the atom is again in its normal state. Consequently, upon viewing the radiation from many different atoms, the entire principal, sharp, and diffuse series and all other lines of the arc spectrum of sodium may be seen.

The *spark* spectrum, appearing when the energy of excitation exceeds 35 volts, is ascribed to ionized atoms in which a *second* atomic electron has been raised to a higher level as the result of a second impact by an electron from the filament. This second atomic electron then causes the emission of radiation as it drops back into lower levels in the ionized atom.

The energy in volts necessary to raise an atom from its normal state into a given excited state is known as the *excitation potential* for that state. The first observation of an excitation potential was made by Franck and G. Hertz in 1914.¹ They showed that electrons with an energy of 4.9 volts or more, but not less than 4.9 volts, would excite the resonance line at 2,536 Å in mercury vapor. They made this observation in the course of a long series of experiments on the energy gained by an electron when passing through a gas under the influence of an electric field. Observations of excitation potentials were important at that time because they furnished direct experimental proof of the existence of those quantized energy levels whose existence had been postulated by Bohr.

109. Absorption and Re-emission of Radiation. (a) *Absorption.*—On the classical theory, the characteristic frequencies emitted by an atom should be identical with the natural frequencies of the atom. Conversely, if light of one of these frequencies falls upon the atom, it should set the atom into resonance, some of the energy being thereby abstracted or “absorbed” from the incident beam. It follows that the *absorption spectrum* of a gas should be exactly the same as its emission spectrum.

For certain spectral lines this conclusion is substantiated by experiment in the well-known phenomenon of the *reversal* of spectral lines, of which phenomena the most conspicuous are the dark lines in the solar spectrum. The reversal of the *D* lines of sodium is a familiar laboratory or lecture demonstration. But many of the lines emitted by a gas in a discharge tube are not ordinarily observable in its absorption spectrum. All absorption lines are found to occur as emission lines, but not the reverse. For example, neon gas is highly transparent to visible light; it does not absorb the red wave lengths whose emission is responsible for the brilliance of neon signs.

The explanation of these facts on the Bohr theory is easily seen. The process of absorption will be just the reverse of the process of emission; a quantum of radiant energy is absorbed while simultaneously the atom jumps from one quantum state into a state of *higher* energy. Thus, in absorption the initial and final states exchange

¹ FRANCK and HERTZ, *D. phys. Ges., Verhandlungen*, vol. 16, p. 512 (1914).

roles. The *possible* frequencies for the lines should, therefore, be the same in absorption as in emission. But in order to absorb a given line, the atom must already be in the state which is the lower state for that line. Now the atoms of a monatomic gas under ordinary conditions are in their normal states of lowest energy. Hence, they can absorb only those lines which start, in absorption, from the normal state. The restriction of the absorption spectrum to certain lines is thus accounted for. For neon, all lines that involve the normal state lie in the ultraviolet. Sodium vapor shows in absorption only the *principal series* [cf. diagram in Fig. 70(b) in Sec. 107]. From the head of the principal series, however, a *continuous* absorption band extends toward higher frequencies. This is obviously due to processes in which the energy of the incident photon $h\nu$ exceeds the ionization potential, so that the atom becomes ionized with the ejection of an electron at considerable speed. That is, we have here photoelectric emission from the sodium atoms, the electron being ejected with kinetic energy

$$\frac{1}{2}mv^2 = h\nu - eV_0,$$

where V_0 is the ionization potential of the atom (and e the numerical electronic charge). The equation is a special case of Einstein's photoelectric equation.

When atoms *in excited states* are present, as in a discharge tube, additional lines may appear in the absorption spectrum.

The energy given to an atom in an absorption process may be lost by the atom subsequently in any one of several ways. Collisions with other molecules may cause the excited atom to return to its original quantum state, the energy of excitation being added to the kinetic energy of motion of the colliding particles. In this case the energy absorbed from the incident radiant energy is converted into heat. Collisions of this sort have been called "collisions of the second kind," the converse collisions, in which kinetic energy is converted into energy of excitation, being called collisions "of the first kind."

(b) *Resonance Radiation and Fluorescence*.—A second way in which an atom may lose an absorbed quantum of energy is by jumping into some other quantum state with the emission of a quantum of radiation. One possibility, which we shall discuss first, is that the atom may drop back into its normal state, from which it was removed by the process of absorption; in this latter case the radiation emitted will have the same frequency as the radiation originally absorbed. The net effect of such absorptions and re-emissions is a powerful scattering of the incident light in all directions. To this phenomenon R. W. Wood gave the name *resonance radiation*.

Wood showed¹ many years ago that a bulb containing sodium vapor at very low pressure would, when irradiated by light from an intense sodium flame, emit the *D* lines, and those only. The resonance was very sharp. An examination of the re-emitted *D* lines showed that they were very narrow, their width corresponding almost exactly to the width predicted by the Doppler effect at the temperature of the sodium vapor. The width of the lines of the exciting source was much greater, since the temperature of the source was much higher than that of the vapor. On analyzing the light which had passed through the vapor, it was found that there was a narrow absorption line at the centers of the broad *D* lines.

This experiment of Wood's was in beautiful agreement with the classical theory. But (the present) Lord Rayleigh showed² that if sodium vapor is illuminated by the *second* line of the principal series, $\lambda = 3,303$ Å, both that line and the *D* lines were emitted by the vapor. The emission of the *D* lines under such circumstances is a case of *fluorescence*, and, like fluorescence in general, is difficult to explain on the classical theory, whereas the explanation by the quantum theory is easy. The absorption of $\lambda = 3,303$ leaves the atom in the excited state known as 4^2P [see Fig. 70(b) in Sec. 107]. From this state it cannot pass at once, for a reason described later (Sec. 142) into the 3^2P state, which is the initial state for the *D* lines, but it can jump into a 4^2S state, in a transition not shown on the diagram, with the emission of an infrared quantum, and then from this state it can jump into the 3^2P state, with the emission of a quantum of $\lambda = 11,404$ or $11,382$, which is the first (doublet) line of the sharp series. Thus, the single quantum absorbed from the incident light is eventually re-emitted as three fluorescent quanta. A second but less likely possibility is a jump from 4^2P to 3^2D , and then into 3^2P with emission of the first line of the diffuse series.

Excitation by the absorption of quanta is quite analogous to excitation by electron impact, except for this one important difference. The atom may become excited when colliding with an electron the energy of which *equals or exceeds* that required for the increase in energy level; whereas excitation by absorption takes place only when the energy of the incident quantum is *exactly equal to* that required to produce the particular change of state.

(c) *Excitation by Collision with Other Atoms.*—The atoms of a gas are continually interchanging energy by collisions due to thermal agitation. As we have seen in Sec. 85, the mean energy of thermal

¹ See R. W. Wood, "Physical Optics."

² STRUTT, R. J., Bakerian Lecture, *Roy. Soc., Proc.*, vol. 98, p. 272 (1916).

agitation of a monatomic gas molecule at 300°K. is of the order of 0.062×10^{-12} erg (*i.e.*, $\frac{3}{2}kT$). This is only a fraction of the energy, 3.36×10^{-12} erg, required to raise the sodium atom from its normal state to the next higher energy level. Very rarely could it happen that a molecule of sodium vapor at room temperature would, as a result of a collision with another molecule, acquire sufficient energy to excite it. If, however, the temperature is raised, as by putting sodium in the Bunsen flame, the average kinetic energy of translation may become large enough so that *collisions between atoms* can raise an appreciable number of atoms to an excited state, and detectable radiation may then result.¹ With further increases in temperature, the higher members of the principal series and members of other series should appear. This is confirmed by experiment.

There is ample evidence to confirm the view that when an atom *A* in an excited state collides with an atom *B* in its normal state, both atoms may undergo a quantum jump. Thus *B* becomes excited, while *A* returns to its normal state, or drops into some intermediate state; and thereafter *B* may radiate one of its own characteristic frequencies. At ordinary temperatures the kinetic energy of the atom is negligible; hence, the energy gained by *B* in such a process cannot appreciably exceed that which is available in the initial excitation of *A*.

To illustrate phenomena involving the transfer of the energy of an excited state from one atom to another, mention may be made of the experiments on mercury vapor² excited to resonance radiation by the absorption of its own line $\lambda = 2,536 \text{ \AA}$ ($h\nu = 4.9$ volts). If with the mercury vapor is mixed the vapor of thallium, the characteristic lines of the latter element appear *in addition to the mercury resonance radiation* when the mixture is illuminated with the *mercury line* $\lambda = 2,536$, the *mercury resonance radiation* being then *weaker than when no thallium is present*. Thallium vapor alone is not excited to resonance by the mercury line. This is explained by assuming that mercury atoms are first raised to an excited state by absorption of the line $\lambda = 2,536$; some of these excited atoms by collision with thallium atoms then transfer their energy of excitation to the latter, which subsequently radiate their characteristic lines in returning to the normal state. The thallium lines so produced correspond to energy transfers in the thallium atom of *less* than 4.9 volts.

¹ At 1800°C., $\frac{3}{2}kT = (2,073/288) \times 0.0597 = 0.43 \times 10^{-12}$ erg. The Boltzmann factor (see next section) for the first excited state relative to the initial state will thus be $e^{-\epsilon/kT} = e^{-3.36/0.43} = 0.00046$.

² LORIA, *Phys. Rev.*, vol. 26, p. 573 (1925); CARIO, *Zeit. f. Physik*, vol. 10, p. 185 (1922); and CARIO and FRANCK, *Zeit. f. Physik*, vol. 17, p. 202 (1923).

Cario showed that the presence of argon in mercury vapor materially reduces the intensity of the resonance radiation when the mixture is illuminated by the $\lambda = 2,536$ line, but without exciting any argon radiation. Yet an examination of the light transmitted through the mixture shows that there is no diminution in *absorption* as a result of the presence of the argon. This effect is explained by assuming that collisions of the second kind take place between the excited mercury atoms and the argon atoms and that the *energy of excitation* of the former is transformed into *kinetic energy* of both atoms.

(d) *Electrical Methods of Observing the Excitation of Atoms.*—Referring to Fig. 71, it is readily seen that if the potential differ-

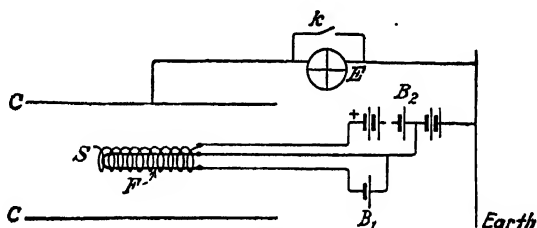


FIG. 72.

ence between grid and filament is sufficiently high, the electrons may be given sufficient energy to ionize by collision the molecules of the vapor in the space between the cylinder and the grid. In Fig. 72, the filament F is so connected to the battery B_2 that the potential difference between F and the grid S is less than, and in the opposite direction to, the potential difference between CC and S . Electrons accelerated from F toward S cannot, therefore, reach CC . The cylinder CC is connected to earth through the sensitive electrometer E and the grounding key k . As long as no positive ions are produced in the space between S and CC , the cylinder should acquire no charge. But when the velocity of the electrons accelerated from F to S becomes sufficient to ionize the vapor, the positive ions thus produced should be attracted to CC and an *ionization current* should be set up which is measurable by the rate at which the electrometer acquires a charge when the grounding key is open.

An effect of this kind was observed by Lenard many years ago. He found that with such gases as air, hydrogen, and carbon dioxide a current, which he thought to be a true ionization current, began to flow when the exciting electrons had fallen through a potential difference of some 11 volts. We have seen, however, from the experiments of Meggers, Foote, and Mohler, that, in the case of sodium vapor, resonance radiation is produced by 2.09-volt electrons and, in the case

of mercury vapor, by 4.9-volt electrons. This resonance radiation, proceeding in all directions, falls on the inside walls of the cylinder, and, if the radiation is of sufficiently short wave length, *photoelectrons* will be expelled from the cylinder. Because of the direction of the field between *CC* and *S*, these photoelectrons will flow toward the grid, and we shall have in the circuit *CB₂S* a *photoelectric current* in exactly the same direction as if there were an ionization current. This effect will begin abruptly as soon as the resonance potential is reached.

To separate the true ionization current from the photoelectric current, Davis and Goucher¹ introduced a coarse wire gauze *G* sur-

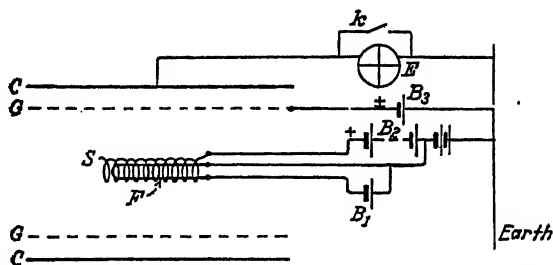


FIG. 73.—The apparatus of Davis and Goucher for studying ionization potentials.

rounding the grid and just inside the plate *CC*. By means of a battery *B₃*,² (Fig. 73), this gauze may be maintained at either a positive or a negative potential with respect to *CC*. When *G* is sufficiently negative with respect to *CC*, photoelectrons ejected from *CC* by the action of resonance radiation will not reach *G* and so cannot escape from *CC*. Photoelectrons ejected from *G*, however, may reach *CC*, and so a photoelectric current may result. If now *B₃* is reversed, so that *G* is positive with respect to *CC*, a photoelectric current in the opposite direction should result, due to photoelectrons ejected from *CC*. In either case the potential difference between *CC* and *G* is kept small as compared with that between *G* and *S*. The true ionization current cannot be reversed by reversing the direction of *B₃*, since the positive ions produced between *S* and *G* acquire sufficient velocity so that, passing through the gauze, they will reach *CC* in spite of any small opposing field between *CC* and *G*.

With this type of apparatus, Davis and Goucher obtained, for mercury vapor, curves of the type shown in Fig. 74, in which the abscissas are the accelerating potentials between filament and grid

¹ *Phys. Rev.*, vol. 10, p. 101 (1917).

² These "batteries" are really variable potentiometric sources of potential difference.

and the ordinates are the currents to the cylinder CC as measured by the rate of charge of the electrometer. Curve A was obtained when the gauze G was *positive* with respect to the cylinder; curve B when the gauze was *negative*. No current was observed until the accelerating voltage reached 4.9 volts. Beyond that voltage a current was observed which could be reversed by reversing the battery B_3 , indicating that the current was a photoelectric current. At point c , corresponding to 10.3 volts, a sudden change in the variation of the currents with increasing potential took place, indicating that the cylinder was beginning to collect positive ions. Up to point c , curve A is qualitatively the mirror image of curve B .

The beginning of the photoelectric currents at 4.9 volts is due to the excitation by electron impacts of the mercury line $\lambda = 2,536$, which then acts photoelectrically on the cylinder and the gauze. The rapid increase in photoelectric current which occurs when the voltage reaches 6.7 volts is attributed to the generation, at that voltage, of the mercury line $\lambda = 1,849$; whereas the sudden break in both curves at 10.3 volts is indicative of the beginning of the real ionization current. The

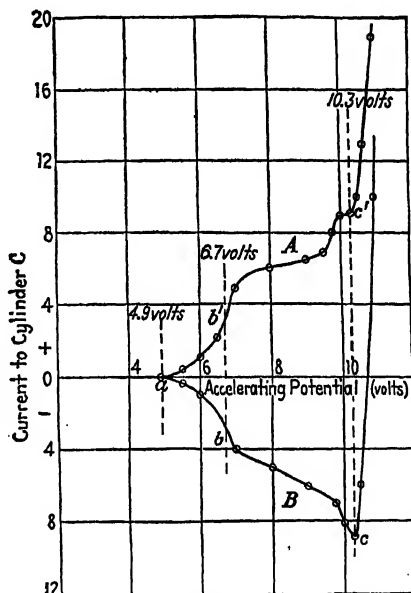


FIG. 74.—The resonance and ionization potentials of mercury vapor, as observed by Davis and Goucher.

quantities 4.9 and 6.7 volts are spoken of as *resonance potentials* of mercury vapor, and 10.3 volts as its *ionization potential*.

(c) *Ionization Potentials*.—As was remarked in Sec. 105, the ionization potential of an atom can also be calculated from the convergence limit of its "principal series," or the series for which the normal state is the final state in the emission of a line. The ionization potential is the equivalent of this convergence limit in volts. Ionization potentials calculated in this way from spectroscopic data agree well with those obtained by electrical methods. Higher ionization potentials can also be defined, the *second* ionization potential representing the energy necessary to remove a *second* electron from the atom; and so on. These higher ionization potentials can be calculated from the convergence limits of other series. Because of the greater

precision obtainable, the spectroscopic method of determining ionizing potentials is usually employed.

For example, in the arc spectrum of helium there is a singlet series of lines due to transitions ending on the normal state, with a convergence frequency¹ $\bar{\nu} = 198,298 \text{ cm.}^{-1}$. According to Eq. (147a) in Sec. 108, this corresponds to a first ionization potential of 24.54 volts. That is to say, it takes energy equal to 24.54 electron-volts to remove 1 of the 2 electrons from a helium atom. When this has been done, the ion behaves as a 1-electron atom. Equation (145) in Sec. 106 gives (with $n_1 = \infty$, $n_2 = 1$), as its ionization energy, $4R$, where R is the Rydberg constant for helium or $R = 109,722.40$. Thus $\nu_{\infty} = 4R$ or $\nu_{\infty} = 438,888$. The second ionization potential of helium is thus $438,888/8,079 = 54.32$ volts.

Values of many of the ionization potentials for the lighter elements are given in Table IV. They have been calculated from data in Bacher and Goudsmit's "Atomic Energy States," the numbers there given for the "absolute value of the lowest state" being divided by 8,079 to obtain modern values for the ionization potential in volts. First ionization potentials (for the neutral atom) are given under "I," second under "II," and so on.

TABLE IV.—SOME IONIZATION POTENTIALS IN VOLTS

Element	I	II	III	IV	V	VI
H 1	13.58					
He 2	24.54	54.33				
Li 3	5.38	75.51	122.2			
Be 4	9.31	18.18	153.6	217.3		
B 5	8.30	25.1	37.87	258.9	339.5	
C 6	11.25	24.34	47.80	64.38	391.1	
N 7	14.52	29.57	47.54	77.3	97.7	
O 8	13.60	35.03	55.04	77.2	114.	137.9
F 9	18.7	34.7				
Ne 10	21.53	40.9	63.3			
Na 11	5.13	47.2				
Mg 12	7.63	15.01	80.			
Al 13	5.98	18.79	28.40			
Si 14	8.14	16.32	33.44	45.07		
P 15	19.9	30.1	64.92	
S 16	10.3	23.4	35.0	47.22	62.1	87.91
Cl 17	13.00	18.36	39.8			
Ar 18	15.73	27.82				
K 19	4.33	31.8	47.0			

¹ BACHER and GOUDSMIT, "Atomic Energy States—as Derived from the Analyses of Optical Spectra," p. 220.

Many beautiful experiments have been devised to extend the data and theories regarding the excitation of atoms by impact or resonance. The reader will find further study in this field, from original sources, very fascinating. The literature is extensive.

110. The Boltzmann Distribution Law.—In the last section, one other easy method of exciting spectral emission has not been mentioned. Many elements in gaseous form, especially the vapors of metals, emit line spectra upon being heated; thus sodium vapor heated by a Bunsen flame emits the familiar *D* lines, whether the vapor is mingled with the flame itself or is enclosed in a tube. In such cases, it is evident that the atoms or molecules are excited by *thermal agitation*.

If the substance is in thermal equilibrium, a simple theoretical formula can be given for the relative numbers of the atoms or molecules that are in each quantum state. It is deduced from statistical mechanics that, in the case of thermal equilibrium, the average number N_i in a quantum state in which the atomic or molecular energy is ϵ_i is

$$N_i = C e^{-\frac{\epsilon_i}{kT}} \quad (148a)$$

where k is Boltzmann's constant (Sec. 85), T is the absolute temperature, and C is a constant of proportionality. If N is the total number of atoms or molecules, we must have

$$N = \sum_i N_i = C \sum_i e^{-\frac{\epsilon_i}{kT}},$$

the sum extending over all possible quantum states. By eliminating C between these two equations we can also write

$$N_i = \frac{N e^{-\frac{\epsilon_i}{kT}}}{\sum_j e^{-\frac{\epsilon_j}{kT}}} \quad (148b)$$

The quantum states referred to here are the fundamental states, to each of which, in wave-mechanical theory, there corresponds an independent wave function. Often, however, a number of states have the same energy, and then it may be more convenient to group these into a multiple state. Suppose we have formed in this manner all the multiple states that we can, no two multiple states having the same energy. Let these multiple states be numbered off in a single series, and let the energy of multiple state number τ be ϵ_τ and the number of fundamental states composing it w_τ ; the number w_τ is

often called the *statistical weight* of the multiple state. Then by Eq. (148a), when thermal equilibrium exists, the number N_i is the same for each of the fundamental states composing a multiple state, so that the number of atoms or molecules in multiple state number τ is just $w_\tau N_i$. Calling this number N_τ , we can write for it, replacing ϵ_i by ϵ_τ in (148a),

$$N_\tau = C w_\tau e^{-\frac{\epsilon_\tau}{kT}}. \quad (148c)$$

These formulas constitute special cases of what is known as the *Boltzmann distribution formula* for quantized systems. We have already encountered an example of the formula in Eq. (119) in Sec. 91 for the distribution of a set of harmonic oscillators in thermal equilibrium. Maxwell's law constitutes an example of the corresponding formula in classical theory [cf. Eq. (112b) in §Sec. 90].

As an example, the sodium *D* lines result, as we have seen, from transitions between either of two excited levels lying close together and the normal level (cf. Sec. 107). If we treat the upper two levels for our purpose as a single composite level, wave-mechanical theory indicates that for it $w = 6$, whereas, for the normal level, $w = 2$. The two levels lie $h\nu = 3.36 \times 10^{-12}$ erg apart. Hence, indicating them by subscripts 1 and 0, respectively, we have for the numbers of atoms in them,

$$N_1 = 6C e^{-\frac{\epsilon_1}{kT}}, \quad N_0 = 2C e^{-\frac{\epsilon_0}{kT}}.$$

In the case of sodium in a Bunsen flame at 1800°C ., this gives

$$\frac{N_1}{N_0} = 3 e^{-(\epsilon_1 - \epsilon_0)/kT} = 3 e^{-3.36 \times 10^{-12} / (2073 \times 1.38 \times 10^{-16})} = 2.3 \times 10^{-5}.$$

Thus only a very small part of the sodium atoms are excited at any given time. They suffice, however, to cause a considerable emission of sodium light.

111. The Extension of Bohr's Theory.—In his original paper, Bohr remarked that the orbit of the electron in a hydrogen atom might be an ellipse instead of a circle. A detailed theory of elliptical orbits was developed by Sommerfeld several years later. The geometrical ideas involved still possess a certain interest.¹

According to the laws governing planetary motion, or any motion under an inverse-square force, an elliptical orbit will have one of its foci at the nucleus, and the energy of the system will depend only

¹ For a more complete account see A. Sommerfeld, "Atomic Structure and Spectral Lines," Methuen & Co., Ltd., London, 1929; also Ruark and Urey, "Atoms, Molecules and Quanta," p. 132.

on the length of the major axis of the ellipse. The orbit will lie in a fixed plane, so that the motion can be described by means of two coordinates, for which we may take polar coordinates r , θ , with the origin at the nucleus. Then, as θ increases through 2π , r increases from its minimum value at one end of the ellipse to a maximum at the other end, after which it decreases again to a minimum. Thus r executes what is called a "libration" during each revolution of the electron. Sommerfeld assumes that the same generalized quantum condition which had been postulated for the harmonic oscillator and the circular orbit, as represented by Eq. (131) in Sec. 102, will hold for r ; i.e., he assumed that

$$\oint p_r dr = n'h,$$

where p_r is the momentum in the radial direction, the integral is taken throughout a libration, and n' is a positive integer or 0. For θ , he assumes with Bohr that the angular momentum is an integer k times $h/2\pi$. (In his books Sommerfeld writes n_ϕ or just n for k and sometimes n_r for n'). Sommerfeld showed then that the energy depends only on the "total" quantum number,

$$n = k + n',$$

and is, in fact, the same function of n as in Bohr's theory [Eq. (136) in Sec. 103]. Thus up to this point nothing is gained except a greater variety in the possible types of orbit. For $n = 1$, we must have $n' = 0$, since k cannot be less than 1; i.e., only the Bohr circle is possible. But, for $n = 2$, we can have either the Bohr circle, with $k = 2$ and $n' = 0$, or an ellipse with major axis equal to the diameter of the circle, corresponding to $k = 1$, $n' = 1$, and so on.

Sommerfeld then showed, however, that, if allowance is made for the known *variation of electronic mass with speed*, the energy of the elliptical motion is slightly different from that of the circular motion. The ellipse is also caused to precess slowly about the nucleus. In this way, he arrived at a splitting or fine structure of the levels of hydrogen, or of any other 1-electron atom. This splitting appeared to be in quantitative agreement with observation. Sommerfeld arrived, in fact, at the same expression for the energy as was obtained later from the wave-mechanical theory as modified to bring it into harmony with relativity. Sommerfeld's theory, however, included no consideration of the effects of electron spin, which was discovered much later; hence, we shall postpone all discussion of the fine structure of the hydrogen lines until we can take it up on a more adequate basis (Sec. 149).

In the meantime, Bohr and others were endeavoring to extend the theory to atoms containing more than 2 electrons. Little quantitative success was achieved in this direction. In particular, no plausible arrangement of orbits could be devised for the 2 atoms of helium which would give the correct value for the second ionization potential (54.3 volts). In a broader way, Bohr endeavored to understand the known variation of chemical and physical properties from atom to atom through the periodic table. His results were so unsatisfactory, however, that two chemists, Lewis and Langmuir, for a time revived the old static atom of J. J. Thomson in a modified form. A striking feature in the periodic table is the tendency for chemical properties to recur in periods of 8 elements, especially near the beginning; it seemed that this feature might be due to some special stability attaching to an arrangement of 8 electrons at the corners of a cube. Nothing beyond qualitative results could be achieved on this theoretical basis, however, without laying down so many postulates that the resulting theory possessed little significance.

The truth is that two essential principles for the understanding of complex atoms were at that time unknown—the Pauli exclusion principle and the phenomenon of electron spin. Accordingly, we shall follow the old quantum theory no further and shall devote the next chapter to a description of the modern theory of wave mechanics.

CHAPTER VII

WAVE MECHANICS

One must be prepared to approach the subject of this chapter philosophically, prepared to accept conclusions which are, at first thought, seemingly at variance with our senses and with a belief that has persisted almost unquestioned from the time of the Greeks, *viz.*, that matter is made up of particles. We have seen in Chap. III that light possesses both undulatory and corpuscular characteristics. But even so, we might say, light differs from matter; whereas we can determine the nature of light only by indirect observation, matter we can *see*. We observe directly that a handful of sand is made up of real particles. By no stretch of imagination do *they* exhibit wavelike characteristics. The particles of sand, which we see so clearly, are, however, made up—so we *think* (!)—of molecules and atoms and electrons and protons and neutrons, *none of which we can see directly any more than we can see light waves or photons*. It is with these *so-called* particles, evidence concerning which is just as indirect as with photons, that the wave theory of matter is primarily concerned.

We shall attempt in this chapter to give merely a brief introduction to this most important subject. Our purpose will be to show how the concept of matter waves can be developed more or less naturally from the previous concepts of both classical and quantum physics and to summarize the pertinent experimental evidence. Certain features and results of the new mathematical theory known as *wave mechanics* will also be described, but no complete presentation of mathematical details will be attempted, as this would require mathematical tools and technique much beyond the scope of this book.

112. Matter Waves. (*a*) *Matter and Energy.*—With the discovery of the law of the conservation of energy toward the middle of the nineteenth century, it became obvious to physicists that the physical universe is made up of two great entities, *viz.*, matter and energy, each of which is “conserved.” These two great conservation laws provided much of the foundation upon which classical physics was built. In large part, the physicists of the latter half of the nineteenth century busied themselves with attempts to determine the fundamental nature and properties of these two entities. By 1900, the

corpuscular nature of matter—molecules, atoms, electrons—had become firmly established; likewise the undulatory nature of light. By 1910, furthermore, Planck's quantum theory and the Einstein photoelectric equation together with various lines of experimental evidence had made it clear that, impossible as it then seemed, physicists would perhaps have to accept the hypothesis that light itself possesses corpuscular characteristics. The decade following 1910 saw the firm establishment of the quantum theory, and by 1920 the dual nature of radiant energy was generally recognized and accepted, even if not understood. With the gradual accumulation of experience, physicists had almost become accustomed (perhaps "resigned" would be more nearly correct!) to using either the corpuscular or the undulatory theory of light, according to the experimental phenomena in hand.

During all this time, there was no suggestion that *matter* was anything but corpuscular. Matter was made up of atoms, which in turn were made up of electrons and nuclei. But, in 1924, Louis de Broglie¹ made the very bold suggestion that particles of matter, and in particular electrons, might possess certain undulatory characteristics, so that they, too, might exhibit a dual nature. He suggested also a way in which the undulatory characteristics of electrons might perhaps furnish a new basis for the quantum theory. He did not develop his ideas into an exact theory, however. His way of thinking was rather akin to that of the ancient Greeks. The reasoning used might almost be paraphrased as follows: "(1) Nature loves symmetry. (2) Therefore the two great entities, matter and energy, must be mutually symmetrical. (3) If (radiant) energy is undulatory and/or corpuscular (?), therefore matter must be corpuscular and/or undulatory (?)."

Had nothing further happened, de Broglie's speculations would doubtless soon have been forgotten. Reasoning of this sort, in the mind of a genius, may *point the way* to an advance, but, before definite scientific progress can be achieved, the new ideas must be precisely formulated and compared with experiment. The advantages of doing this were overlooked by the ancient Greeks. De Broglie's loosely woven speculation, however, set a German, E. Schrödinger, to thinking; and he discovered how to develop certain of de Broglie's ideas into a precise mathematical theory.² With additions made to

¹ DE BROGLIE, *Phil. Mag.*, vol. 47, p. 446 (1924); *Ann. de Physique*, vol. 3, p. 22 (1925).

² SCHRÖDINGER, *Ann. d. Physik*, vol. 79, pp. 361, 489, 734; vol. 80, p. 437; vol. 81, p. 109 (1926).

it by Born, Heisenberg, Dirac, and others, this theory has become the astonishingly successful quantum mechanics of the present day. In what follows, we shall endeavor to arrive at the basic assumptions of Schrödinger's theory by developing de Broglie's ideas in a natural and logical manner.

(b) *Matter Waves*.—In his first paper, de Broglie is concerned primarily with developing a theory of light in terms of "light quanta," or photons. If the energy of the light is concentrated in photons moving like particles, how are the phenomena of interference to be understood? There must be waves of some sort associated with the photons, in order to account for the observed interference effects. We can no longer suppose that the energy is spread out over these waves, as in classical theory; nevertheless, the waves must somehow determine where, in an interference pattern, the photons can produce effects by being absorbed. The details of interference patterns depend largely upon the phase relations of the waves; hence, de Broglie called the latter *phase waves*. He assumed the frequency ν of the waves to be such that the energy in a photon equals $h\nu$.

But why should such waves be associated only with photons? May not material particles have wave properties as well? A material particle carries energy; and "it would seem that the basic idea of the quantum theory is the impossibility of imagining an isolated quantity of energy without associating with it a certain frequency."¹ Material particles ought, therefore, like photons, to be accompanied by phase waves of some sort; and these waves ought, under suitable circumstances, to give rise to interference effects. Furthermore, to complete the analogy, the waves associated with a particle moving at speed v should have a frequency equal to the energy of the particle divided by h .

Regarding the energy of a material particle, however, there is a divergence between the assumptions of de Broglie and of Schrödinger. De Broglie includes in the energy the internal or rest energy, m_0c^2 , the existence of which is suggested by the theory of relativity as the cause of the existence of the rest mass m_0 (Sec. 69). Schrödinger, on the other hand, prefers to construct first a nonrelativistic theory. He assumes the mass of the particle to be constant and equates just the ordinary energy, kinetic plus potential, to $h\nu$. In this respect we shall follow Schrödinger, reserving comment on the relativistic form of the theory, which must be used for high-speed particles, to a later section (Sec. 126). We shall assume, therefore, that any material particle of mass m moving at speed v has associated with it in some manner waves

¹ Translated from *Ann. de Physique*, vol. 3, p. 32 (1925).

of frequency ν given by the relation

$$h\nu = \frac{1}{2}mv^2 + V, \quad (149)$$

where h is Planck's constant, m is the mass of the particle, and V is its potential energy at the moment, due to electrical fields or to other causes.

The physical nature of these waves was left indefinite by de Broglie. We cannot suppose that a material particle *is* just a group of waves; for then its mass and energy, and also its charge if it has a charge, would of necessity be spread out over these waves and would, in consequence, soon become scattered widely in space, contrary to the observed facts. It is of the nature of waves to diverge toward all sides. We should not be disturbed, however, by the impossibility of visualizing the waves. We should remember our experience in optics. Using classical theory, it was easy to picture light as wave motion; but, if we retain this concept, it is very difficult to picture a beam of light as a moving stream of photons. Similarly, as long as we retain the particle concept of an electron or proton, we cannot hope to form a concrete picture of the accompanying waves. Perhaps, even, they are only mathematical waves, so to speak, a device that we employ for the purpose of making calculations and predicting the results of observation. In adopting this standpoint, we are doing only what has been done, in various ways, many times previously in physical science. Strictly speaking, we have no very exact knowledge of the fundamental nature of a magnetic field, and yet we do not hesitate to use the symbol H with all due familiarity. In Sec. 120 will be described more precisely the extent to which it has been found possible to attach physical significance to matter waves.

113. Mechanics as Geometrical Optics of the Waves.—If waves of some sort are associated with all material particles and play a part in determining their motion, as light waves certainly do for the photons, then there should be a parallelism between the laws of mechanics and the laws of wave motion. For the particles certainly obey the laws of mechanics in some cases, at least. Now it was pointed out long ago by Hamilton that there does exist a close parallelism between the laws of mechanics and the laws of ordinary *geometrical optics*. De Broglie suggests, accordingly, that the ordinary laws of particle mechanics may represent an approximation which is valid under such circumstances that the laws of geometrical optics hold for the matter waves.

One of the fundamental laws of ordinary geometrical optics is that, in a homogeneous medium, light travels in straight lines or rays.

This assumption is very nearly correct as long as the lateral dimensions of the beam are large compared with the wave length of the light. Under these conditions one might say that the "particle" characteristics¹ of light appear to predominate, although we now know that the rectilinear propagation of light is entirely consistent with the wave theory. When, however, the cross section of the beam is of the same order of magnitude as the wave length of the light, rectilinear propagation no longer holds, diffraction phenomena are observed, and undulatory characteristics of light predominate. If we carry the similarity over to mechanics, might we not expect that, for very small

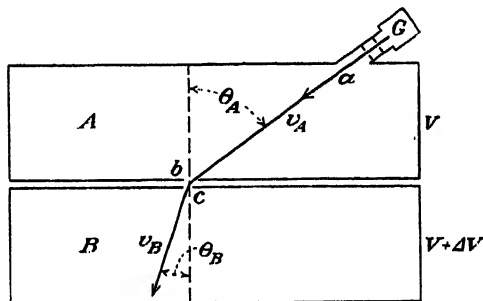


FIG. 75.—Showing "refraction" of a pencil of electrons.

particles of matter, the ordinary laws of mechanics—found by Newton to be applicable to the phenomena of "macromechanics"—would fail, and, by analogy with light, we should find that matter shows undulatory properties in the realm of "micromechanics"?

If we assume that the phenomena of ordinary mechanics constitute those of geometrical optics for matter waves, then, as de Broglie showed, we can find out a great deal about the properties that these waves must have in order that, under suitable conditions, the laws of mechanics may hold. Ordinary geometrical optics might be said to be based upon two laws, the law of reflection and the law of refraction. The law of reflection presents no difficulty, for it is exactly the same as the law governing the rebound of an elastic body from a hard wall. The refraction of the waves, on the other hand, obviously corresponds to the deflection of the particle by the action of forces. It will be worth while to study this latter phenomenon in greater detail.

114. Refraction of Matter Waves.—Let us first analyze a mechanical phenomenon which imitates a simple case of refraction in geometrical optics. Let a pencil of electrons from a suitable gun *G* (Fig. 75),

¹ It will be recalled that, from this observed rectilinear propagation, Newton argued that light is corpuscular, since, like free bodies in motion, it travels in straight lines.

in which they are accelerated by a potential drop of V_0 electrostatic units, enter through orifice a an enclosed metal box A , the potential of which is V . Let these electrons emerge from A through orifice b into the space between box A and box B , the latter of which is maintained at a potential $+\Delta V$ with respect to A . The electric field between the two boxes will give the electrons a component of velocity in a direction at right angles to the adjacent surfaces, and the electrons will enter box B through orifice c in a direction different from that in box A . This change of direction is closely analogous to that experienced by light in passing from one medium to another. Let v_A and v_B (Fig. 76) be the velocities of the electron beam in box A and box B , respectively; and let the corresponding kinetic

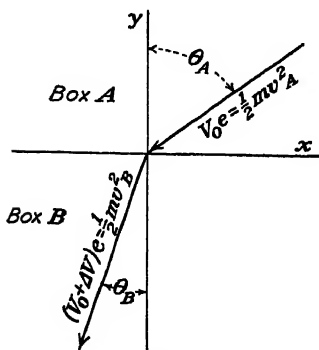


FIG. 76.

energies be

$$\frac{1}{2}mv_A^2 = V_0 e, \quad \frac{1}{2}mv_B^2 = (V_0 + \Delta V)e, \quad (150a,b)$$

e being the numerical charge on the electron. Since the electric field does not change the component of velocity in the x direction, we may write

$$v_A \sin \theta_A = v_B \sin \theta_B, \\ \therefore \frac{v_B}{v_A} = \frac{\sin \theta_A}{\sin \theta_B}, \quad (151)$$

where θ_A and θ_B are the angles which the two beams make with the normal to the adjacent faces of the boxes. From Eqs. (150a,b) we have

$$\frac{v_B^2}{v_A^2} = \frac{(V_0 + \Delta V)e}{V_0 e} = 1 + \frac{\Delta V}{V_0},$$

from which and from Eq. (151) we obtain

$$\frac{\sin \theta_A}{\sin \theta_B} = \frac{v_B}{v_A} = \sqrt{1 + \frac{\Delta V}{V_0}}. \quad (152)$$

Now this result might also be expressed by introducing an equivalent refractive index μ defined by the equation

$$\frac{\sin \theta_A}{\sin \theta_B} = \mu.$$

Then

$$\mu = \sqrt{1 + \frac{\Delta V}{V_0}}. \quad (153)$$

We see, therefore, that, for any given pair of values of V_0 and ΔV , the bending of the electron beam can be described by saying that the space B has a certain "electron index of refraction" with respect to space A . Or, we can proceed as is done in optics and introduce an index of refraction for each region. Let us assume as indices of refraction for electrons in spaces A and B , respectively,

$$\mu_A = k' \sqrt{V_0}, \quad \mu_B = k' \sqrt{V_0 + \Delta V},$$

in which k' is an arbitrary constant. Then, as in optics,

$$\mu = \frac{\mu_B}{\mu_A} = \frac{\sqrt{V_0 + \Delta V}}{\sqrt{V_0}} = \sqrt{1 + \frac{\Delta V}{V_0}}.$$

Clearly we can account for the motion of the electron along any path by assuming that, at any point where its potential is V , its refractive index is

$$\mu = k' \sqrt{V}.$$

It is important for the validity of this result that the arbitrary additive constant in the potential V be so chosen that $V = 0$ at a point where the kinetic energy of the electron is zero. The kinetic energy at any other point is then

$$\frac{1}{2}mv^2 = eV,$$

e being the numerical charge on the electron. If we eliminate V between the last two equations, we obtain

$$\mu = k''v,$$

where $k'' = k' \sqrt{m/2e}$ and represents a new arbitrary constant.

In this latter form, the result is applicable to the motion of any sort of particle in a field of force. For we could set up an electric field in which at each point the potential energy of an electron, $-eV$, would be the same as the potential energy of the other particle in its own force field; then a particle and an electron, projected with equal speed and in the same direction from corresponding points, would traverse similar trajectories. Thus *the refractive index necessary to account for the motion of a particle in a force field varies from point to point along the path in proportion to the velocity of the particle.*

The refractive index for waves, on the other hand, is inversely proportional to the wave speed u :

$$\mu = \frac{k'''}{u}.$$

This rule is a direct inference from Huygens' principle, which must hold for anything that we may call waves. Hence, if the waves, moving according to the laws of geometrical optics, are to follow the particle along its path, we must have at every point

$$u = \frac{b}{v} \quad (154)$$

in terms of a new constant of proportionality b , or the *wave speed must be inversely proportional to the particle speed*. Or, we can also express this result in terms of the constant total energy E and the (perhaps variable) potential energy U of the particle. For

$$\frac{1}{2}mv^2 = E - U = h\nu - U,$$

if we introduce the assumption that the frequency ν of the waves is given by the Einstein relation, $E = h\nu$, h being Planck's constant. Hence

$$v = \left[\frac{2(h\nu - U)}{m} \right]^{1/2}, \quad u = b_1(h\nu - U)^{-1/2}, \quad (155a, b)$$

$$\mu = \mu_0 \sqrt{h\nu - U}, \quad (156)$$

b_1 and μ_0 being new arbitrary constants ($b_1 = b\sqrt{m/2}$, $\mu_0 = k'''/b_1$).

According to this last equation, in which U is a function of position in space, the refractive index for the waves must be supposed to vary from point to point, in somewhat the same way as it does for light waves in a nonhomogeneous piece of glass. At any given point, furthermore, μ varies with ν , *i.e.*, the waves exhibit the phenomenon of dispersion. This is true even in free space, where $U = 0$.

Parenthetically, it should be remarked, to avoid misunderstanding, that the "electron microscope," of which the student may have heard, does not depend for its operation upon the wave properties of the electrons. The electrons in this microscope move according to classical mechanics, although their motion is sometimes treated in terms of the equivalent refractive index described above. The only connection matter waves have with the electron microscope is that their wave length would ultimately set a limit to the possible resolving power that can be attained, just as the resolving power of ordinary microscopes is limited by the finite wave length of light. The wave length of the electron waves, however, we shall presently find to be so short that at present the resolving power of the electron microscope is limited by other factors.

115. Fermat's Principle and the Principle of Least Action.—For the benefit of those students who may be interested in the more

abstract formulations of mechanics, an alternative and more complete treatment of the refraction of matter waves will also be given, along lines suggested by de Broglie.

One way of stating the law of the propagation of light in geometrical optics is in terms of Fermat's *principle of least time*. This principle is a direct deduction from Huygens' principle and states that, of all possible paths which a ray of light might conceivably take between two given points, it chooses that path which makes the time required

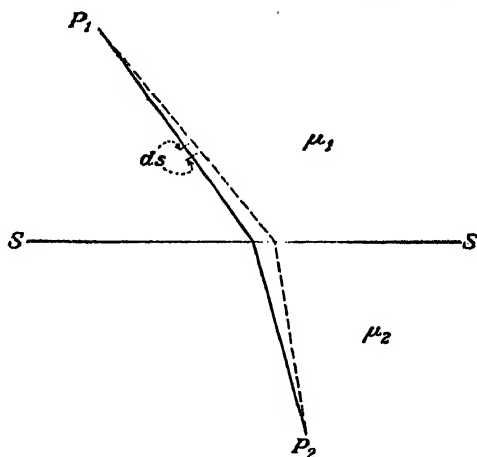


FIG. 77.—The solid line P_1P_2 represents the path taken by a beam of light; the dotted line represents a geometrically possible near-by path.

a minimum (more exactly, an extremum, a minimum if the points are not too far apart; otherwise, often a maximum). The principle is equally applicable to the propagation of waves of any kind.

To express the principle in mathematical form, consider a surface SS (Fig. 77) separating two mediums of respective indices of refraction μ_1 and μ_2 . Let the full line represent the path actually taken by a monochromatic ray of light of frequency ν in passing from any point P_1 to any other point P_2 , the dotted line being a near-by geometrical path. Let ds represent an element of the path, and u the velocity of light of frequency ν in the medium in which ds is located. Then the time τ required for the light to travel from P_1 to P_2 is

$$\tau = \int_{P_1}^{P_2} \frac{ds}{u}.$$

Fermat's principle of least time is stated by the equation

$$\delta\tau = \delta \int_{P_1}^{P_2} \frac{ds}{u} = 0, \quad (157)$$

where $\delta\tau$ means "the difference between the value of τ for the natural path and for an infinitely near path." An alternative form of the last equation is obtained by multiplying Eq. (157) by c , the velocity of light in free space, and introducing the refractive index, $\mu = c/u$:

$$\delta \int_{P_1}^{P_2} \frac{c}{u} ds = \delta \int_{P_1}^{P_2} \mu ds = 0. \quad (158)$$

Now in form these equations are the same as those expressing the *principle of least action* in mechanics, which describes, or determines, the "natural" path taken by a body, regarded as a material particle, in its passage from one point to another under the action of conservative forces. Let the line AB (Fig. 78) represent the natural path of, say a projectile of mass m , air resistance being eliminated. Let ds be an element of length of the path and let v be the velocity of the projectile. Then the "action" A between any two points P_1 and P_2 on the natural path is defined as the line integral of the momentum over the path between those points. That is,

$$A_{P_1}^{P_2} = \int_{P_1}^{P_2} mv ds.$$

The principle of least action now states that, of all imaginable paths between P_1 and P_2 , the projectile chooses the path which makes A a minimum (or at least an extremum) as compared with the value which A would have for any neighboring path that *might* be followed from P_1 to P_2 . It must be assumed that all paths are traversed with the same total energy. Or, in mathematical form,

$$\delta A = \delta \int_{P_1}^{P_2} mv ds = 0, \quad (159)$$

where δA has a meaning analogous to that of $\delta\tau$ in Eq. (157). Another form is obtained if we substitute for v in terms of the total energy E and the potential energy V ; since then $\frac{1}{2}mv^2 = E - V$,

$$mv = \sqrt{2m(E - V)},$$

and (159) can be written

$$\delta \int_{P_1}^{P_2} \sqrt{2m(E - V)} ds = 0. \quad (160)$$

In this latter equation, E is to be kept constant as the path is varied, whereas V depends only on position in space.

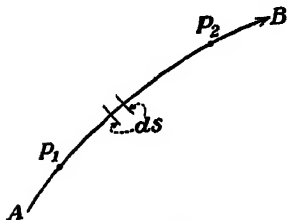


FIG. 78.

These equations for the projectile become the same as those obtained previously for light waves if we suppose that

$$\frac{1}{u} = \frac{v}{b} = \frac{\sqrt{2m(E - V)}}{bm} \quad (161)$$

or

$$\mu = \frac{c}{u} = \mu'_0 \sqrt{2m(E - V)}, \quad (162)$$

where $\mu'_0 = c/(bm)$; with these substitutions, (157) becomes the same as (159), and (158) the same as (160), except for constant factors (m , b , μ'_0) which can be canceled out. Thus a ray of light can follow the same path as the projectile, provided the index of refraction varies along the path as given by Eq. (162). If we replace light waves by matter waves, we obtain Eqs. (154) and (156) above, in which U is written for V and the frequency of the waves is such that $E = h\nu$; the arbitrary constant μ_0 in (156) replaces $\mu'_0 \sqrt{2m}$ in (162).

116. The de Broglie Wave Length.—The most significant feature of matter waves will undoubtedly be their *wave length*. It is this property of waves in general that controls the phenomena of diffraction and interference. Even in ordinary *wave optics*, the fundamental entity is the wave length; the frequency ν , which appears in the expression $h\nu$ for the energy of a photon, the fundamental entity in *corpuscular optics*, is actually calculated from measured values of the wave length.

The wave length to be expected for matter waves can be inferred by means of another suggestion made by de Broglie. If we make the right assumption concerning the velocity or refractive index of matter waves, as we have seen, they can be caused to follow the same trajectory as does the particle with which they are associated, in so far as classical mechanics is valid for the particle and the laws of geometrical optics for the waves. But it follows from Eq. (154) that the waves will not actually keep up with the particle, since in general they move with different velocity. It is well known, however, that, whenever the wave velocity varies with the frequency (*i.e.*, whenever dispersion exists), a finite train or group of waves moves with a velocity different from that of the individual waves. This phenomenon is easily observed on water. Close inspection of a group of waves advancing over a water surface will show that the individual waves advance about twice as fast as the group as a whole; new waves continually arise at the rear of the group, pass through it, and die out at the front. In the case of matter waves, de Broglie showed that their group velocity could be made equal to the velocity of the associ-

ated particle, so that a group of them is able to accompany the particle in its motion.

It is readily shown¹ that, if u is the *wave velocity*, a finite group of waves having almost the form of sine waves with wave length λ advances with a *group velocity* u' of magnitude

$$u' = u - \lambda \frac{\partial u}{\partial \lambda}. \quad (163)$$

Here a partial derivative of u is written because u may vary with the refractive index from point to point as well as with λ . If there is no dispersion, as in the case of sound or of light *in vacuo*, $\partial u / \partial \lambda = 0$ and $u' = u$. In the presence of dispersion, however, the group velocity and the wave velocity are not equal. In the latter case, the group velocity is obviously the velocity at which the energy travels; the individual waves function essentially as "phase waves," serving to determine, for example, where the energy shows up in an interference pattern.

Let us now postulate the equality of group velocity and particle velocity for matter waves and see to what conclusions this postulate leads. Then, replacing u' in (163) by the particle velocity v , we have

$$v = u - \lambda \frac{\partial u}{\partial \lambda}. \quad (164a)$$

But, according to Eq. (154), along any path, u must vary in such a way that

$$v = \frac{b}{u}, \quad (164b)$$

b being a constant along the given path. Furthermore, for any type of waves, $u = \nu \lambda$; and the frequency ν of matter waves we have assumed to be such that

$$h\nu = E = \frac{1}{2}mv^2 + U, \quad (165)$$

U being the potential energy and E the total energy of the particle. Differentiating this latter equation with respect to λ , we have²

$$h \frac{\partial \nu}{\partial \lambda} = mv \frac{\partial v}{\partial \lambda}.$$

But substitution of $u = \nu \lambda$ in Eq. (164a) gives

$$v = \nu \lambda - \lambda \frac{\partial}{\partial \lambda}(\nu \lambda) = -\lambda^2 \frac{\partial \nu}{\partial \lambda}.$$

¹ See, e.g., Wood, R. W., "Physical Optics," or Houston, R. A., "Treatise on Light."

² U contains as variables x, y, z ; these are to be kept constant in the differentiation.

Elimination of $\partial v / \partial \lambda$ between these last two equations gives

$$v = -\frac{\lambda^2 m v}{h} \frac{\partial v}{\partial \lambda}; \quad \therefore \frac{\partial v}{\partial \lambda} = -\frac{h}{m \lambda^2}.$$

Integrating, we have

$$v = \frac{h}{m \lambda} + \text{const.}$$

If we set the constant of integration equal to zero in this last equation, we have the product $m v \lambda$ equal to a universal constant. Then along any path, since $\lambda = u/v$, $m v u/v$ is constant, and, since m and v are constant, $v u$ is also constant, which is in agreement with Eq. (164b). It can be shown that no other assumption concerning the constant of integration in the last equation can be harmonized with Eq. (164b). Hence, we can write the last equation in the form

$$\lambda = \frac{h}{p}, \quad p = m v. \quad (166a, b)$$

Here p represents the ordinary momentum of the particle.

Equations (165) and (166a, b) fix the frequency ν and the wave length λ of the waves in terms of the energy E and momentum p of the particle with which they are associated. The wave velocity, if wanted, can then be calculated as $u = \nu \lambda$. Wave lengths given by Eq. (166a, b) are known as *de Broglie wave lengths*. It may be remarked that Eq. (166a) holds for photons as well as for material particles; for the momentum of a photon is $h\nu/c = h/\lambda$ [cf. Eq. (25c) in Sec. 35]. Equation (166a), it should be noted, combines the corpuscular and the undulatory concepts of light in a very intimate way; for λ has a clean-cut meaning only in connection with a *wave* theory, and p , the momentum, is most naturally associated with a moving *particle*.

We can now compute the wave lengths to be expected for electron waves, atom waves, or molecule waves. For an electron moving at a velocity much below that of light with kinetic energy equal to V electron-volts (Sec. 105), we have

$$\frac{1}{2} m v^2 = \frac{eV}{300}, \quad p = m v = \sqrt{\frac{2meV}{300}},$$

e being the numerical electronic charge in electrostatic units. Hence by (166a) its wave length is¹

¹ If V exceeds a few thousand volts, it is necessary to take account of the variation of mass with velocity. Using the relativistic equation for kinetic energy, Eq.

$$\lambda_e = h \sqrt{\frac{150}{meV}} = \frac{12.24}{\sqrt{V}} \times 10^{-8} \text{ cm.} = \frac{12.24}{\sqrt{V}} \text{ \AA}, \quad (167)$$

if we introduce $m = 0.9107 \times 10^{-27}$ gram, $e = 4.803 \times 10^{-10}$ e.s.u. For 100-volt electrons, $\lambda_e = 1.22$ angstroms; for 10,000-volt electrons, $\lambda_e = 0.122$ angstrom.

Wave lengths can be calculated in a similar way for molecules or, for that matter, even for large masses such as billiard balls. The larger the mass, the shorter is the wave length at given speed.

It would be logical now, following the historical order, to describe at once the mathematical theory of matter waves, as worked out by Schrödinger. All that we need, however, for a discussion of experi-

(68) in Sec. 68, we have

$$\frac{eV}{300} = m_0 c^2 \left[\left(1 - \frac{v^2}{c^2} \right)^{-\frac{1}{2}} - 1 \right],$$

c being the speed of light. From this equation, if we write

$$\frac{eV}{300 m_0 c^2} = \rho,$$

we find

$$\left(1 - \frac{v^2}{c^2} \right)^{-\frac{1}{2}} = 1 + \rho, \quad v = c \sqrt{1 - (1 + \rho)^{-2}}.$$

Substituting these values in the relativistic equation for momentum, Eq. (67) in Sec. 67, we have for the momentum

$$p = \frac{m_0 v}{(1 - v^2/c^2)^{\frac{1}{2}}} = m_0 c \sqrt{(1 + \rho)^2 - 1} = m_0 c \sqrt{\rho(2 + \rho)}.$$

Hence, by (166a), which is readily shown to hold in the relativistic case also, the electronic wave length is

$$\lambda_e = h \sqrt{\frac{150}{m_0 e V}} \left(1 + \frac{1}{2} \rho \right)^{-\frac{1}{2}}.$$

Here ρ represents the ratio of the electron's kinetic energy to its relativistic rest energy $m_0 c^2$, and the parentheses represent the effect of the variation in mass on λ_e . For $\rho \ll 2$, i.e., $V \ll 10^6$ volts, we can write

$$\lambda_e = h \sqrt{\frac{150}{m_0 e V}} \left(1 - \frac{\rho}{4} + \frac{3}{32} \rho^2 \cdots \right).$$

The correction to formula (167) is about 2.5 percent for 50,000-volt electrons and about 8.5 percent for 200,000-volt electrons. For $\rho \gg 2$, i.e., $V \gg 10^6$, we can use the expansion in powers of $1/\rho$:

$$\lambda_e = \frac{300 ch}{eV} \left(1 - \frac{1}{\rho} + \frac{3}{2\rho^2} \cdots \right)$$

ments on the diffraction and interference of these waves is their wave length. It may relieve the tedium of so much abstract discussion if we describe next the experiments which have shown that electrons and even molecules actually do possess certain wave properties.

117. Experiments on Electron Waves. (a) *Reflection from a Crystal.*—The first experiments on electron diffraction were reported¹ by Davisson and Germer 3 years after de Broglie's first paper appeared. Davisson and Germer were studying the reflection of electrons from a nickel target and accidentally subjected a target to such heat treatment that it was transformed into a group of large crystals. Anomalies then appeared in the reflection from it. Following up this lead,

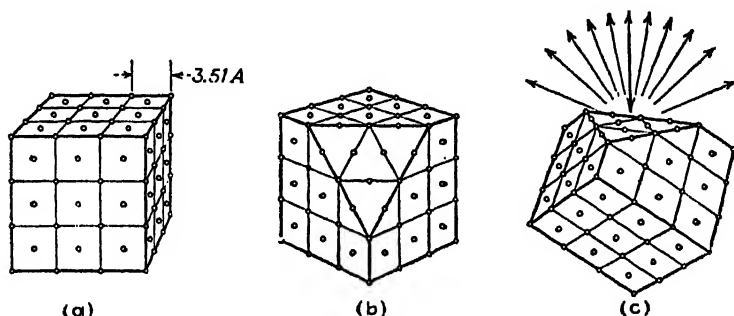


FIG. 79.—(a) A nickel crystal, showing face-centered cubic structure. (b) Same, with a face cut at right angles to a diagonal. (c) An incident beam of electrons (heavy arrow) is scattered in all backward directions.

they prepared a target consisting of a single crystal of nickel and bombarded its surface at normal incidence by a narrow pencil of low-voltage electrons; by means of a suitable "collector" of small aperture, they studied the distribution in angle of the electrons reflected from the crystal. In this reflected beam, they found striking maxima and minima, which they were able to explain in terms of diffraction of the electron waves.

The diffraction of such waves by a crystal is very similar to the diffraction of X-rays, which has been studied extensively, and is described at length in a later chapter (cf. Sec. 186). Here we shall discuss the structure of crystals, and the diffraction of waves by them, only as far as is necessary in order to understand the significance of the experiments on matter waves.

Crystallographic studies by means of X-rays show that the nickel crystal is of the "face-centered cubic" type, as shown in Fig. 79,² i.e., the crystal can be imagined to be constructed of cubical unit cells

¹ DAVISSON and GERMER, *Phys. Rev.*, vol. 30, p. 705 (1927).

² This and certain other figures are used by permission of Dr. Davisson.

each having an atom at each corner and one in the center of each face, with none inside the cube. The atoms are indicated by circles in the figure, certain ones being joined by lines in order to outline the unit cells. The length a_0 of the edge of the unit cube is 3.51 angstroms. Figure 79(b) shows a face cut on the crystal at right angles to one of the diagonals of the cube. In Fig. 79(c) the direction of the incident electrons is shown by the heavy arrow. The lighter arrows show,

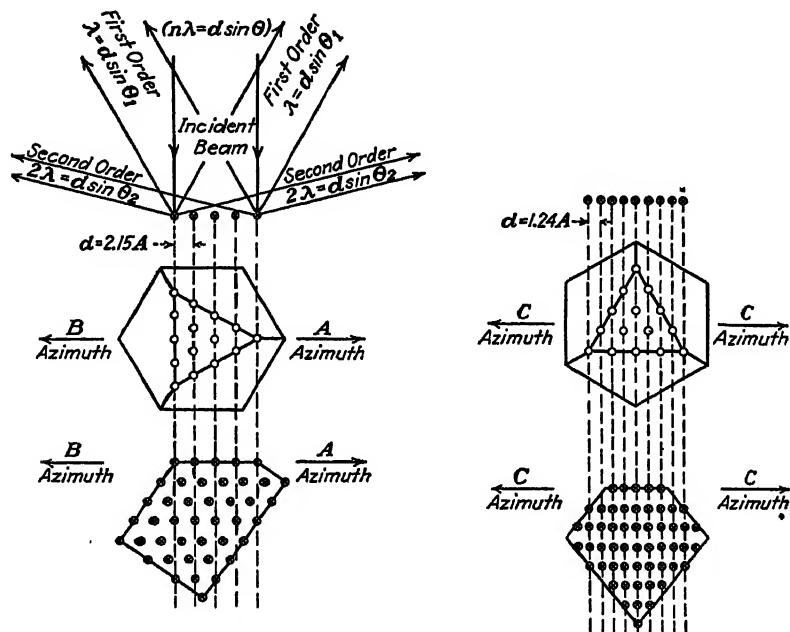


FIG. 80.—Schematic representation of reflection of electron beams from a nickel crystal.

schematically, that these electrons are reflected in all (backward) directions. Considering only the surface layer of atoms in the triangular face of Fig. 79(c), it is readily seen that these atoms are arranged in rows parallel to one side of the triangle, and it is easily shown that the distance d between these rows is

$$d = \sqrt{\frac{3}{8}} a_0 = 2.15 \text{ angstroms.}$$

Now it is fairly obvious that we may regard these rows of atoms as equivalent to the lines of a plane grating of grating space $d = 2.15$ angstroms. Radiation of wave length λ incident normally on such a grating, the plane of incidence being taken normal to one side of the triangle, should be diffracted, as is light from a reflection grating.

according to the well-known law

$$n\lambda = d \sin \theta, \quad (168)$$

where θ is the angle between the (normally) incident beam and the diffracted beam and n is the order of diffraction [Sec. 186(b)]. These relations are shown schematically in Fig. 80. It will be observed, from Fig. 79(c) and from Fig. 80, that the crystal is in reality equivalent to a series of such plane gratings, piled one above the other and separated by a distance $b = a_0/\sqrt{3}$. Radiation penetrating to and diffracted backward from any one of these underlying layers will be combined with that diffracted from other layers, with the result

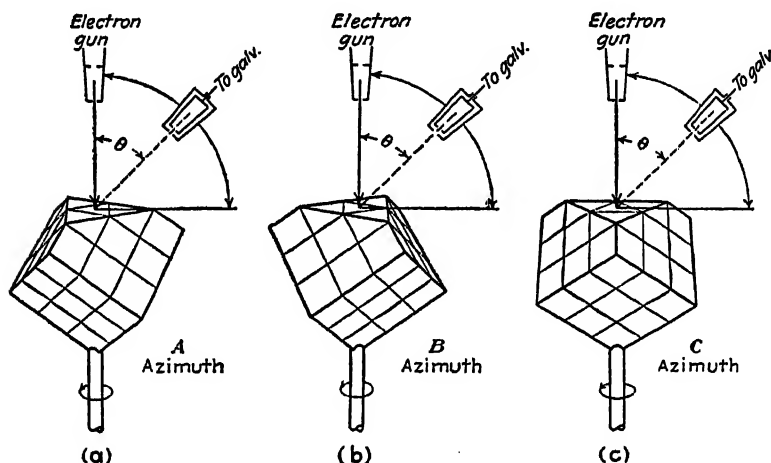


FIG. 81.—Schematic representation of apparatus used by Davisson and Germer in their discovery of the diffraction of electrons.

that, for a given wave length λ of the (normally) incident radiation, we shall expect to find a diffracted beam at the angle θ defined by Eq. (168), unless the beams from the several layers destructively interfere. If conditions are such that destructive interference takes place, the beam to be expected at angle θ will be wholly or partially suppressed.

The apparatus used by Davisson and Germer is represented diagrammatically in Fig. 81. The "electron gun" consists of a heated filament emitting electrons, an accelerating field to give the electrons any desired velocity, and a series of collimating apertures to produce a (nearly) unidirectional beam. This monovelocity beam of electrons strikes the surface of the crystal at normal incidence, and the electrons are reflected, or scattered, in all directions. A collector for measuring the reflected electrons is so arranged that it

can be adjusted to any angular position with respect to the crystal, and thus the intensity of the electron beam reflected within a small solid angle $d\omega$ may be determined. The collector has two walls insulated from each other. A retarding potential is applied between the inner and the outer wall so that only the fastest electrons—those possessing nearly the incident velocity—may enter the inner chamber and be measured by the galvanometer. The crystal may be turned about an axis parallel to the axis of the incident beam, and thus any azimuth of the crystal may be presented to the plane defined by the incident beam and the beam entering the collector. In what follows we shall be concerned mainly with the A and the B azimuths.

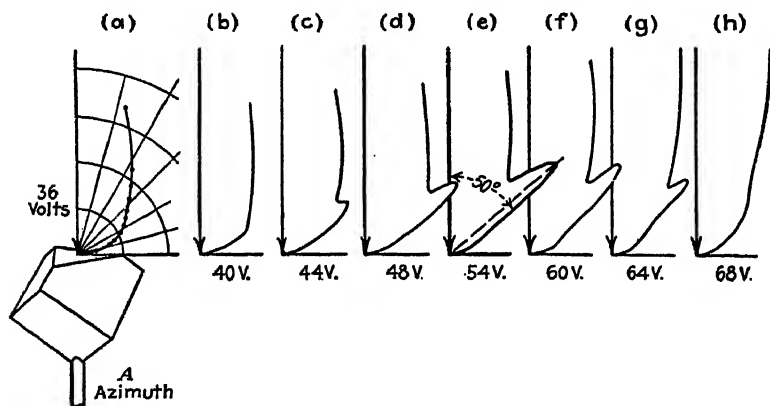


FIG. 82.—The development of the diffraction beam in the A azimuth, showing maximum length of "spur" for 54-volt electrons, at colatitude 50° .

If a beam of low-voltage electrons is incident on the crystal, turned at any arbitrary azimuth, and the distribution of the scattered beam is measured as a function of the colatitude—the angle between the incident beam and the beam entering the collector—a curve similar to that in Fig. 82(a) is obtained, which refers to incident 36-volt electrons. If now the crystal is turned to the A azimuth, the distribution curve for 40-volt electrons [Fig. 82(b)] shows a slight "hump" at about colatitude 60° . With increasing voltage this hump moves upward and develops into a spur which becomes most prominent at 54 volts [Fig. 82(e)], at which voltage the colatitude of the spur is 50° . At higher voltages the spur gradually disappears.

The spur in its most prominent state of development offers convincing evidence for the existence of electron waves. According to Eq. (167), the wave length λ_e of a beam of 54-volt electrons should be

$$\lambda_e = \frac{12.24}{\sqrt{54}} = 1.67 \text{ angstroms.}$$

Since a diffracted beam is observed at 50° , it is obvious that, irrespective of how the underlying layers have cooperated, constructive interference has occurred among the beams diffracted from the several plane gratings; hence, we may use Eq. (168) to determine the wave length. We find

$$\lambda_e = d \sin \theta = 2.15 \sin 50^\circ = 1.65 \text{ angstroms.}$$

This observed value of λ_e agrees excellently with that computed above from de Broglie's formula.

A similar agreement was found for other beams. In the *B* azimuth, a spur was strongly developed¹ with 65-volt electrons at a colatitude 44° . The wave length of 65-volt electrons is found in a similar way to be 1.52 Å, whereas at 44° we have $\lambda_e = 2.15 \sin 44^\circ = 1.49$ Å. In the *A* azimuth, a spur was also found at 55° colatitude with 181-volt electrons. This was interpreted as a second-order beam. For 181-volt electrons, $\lambda_e = 0.91$ Å, whereas second-order diffraction gives, for colatitude 55° ,

$$2\lambda_e = 2.15 \sin 55^\circ, \quad \therefore \lambda_e = 0.88 \text{ Å.}$$

In the *C* azimuth,² for which $d = 1.24$ Å, a strong beam was found at 56° colatitude for 143-volt electrons. The wave length corresponding to 143-volt electrons is 1.02 Å, whereas $1.24 \times \sin 56^\circ = 1.03$ Å—again an excellent agreement. Over 20 such beams were reported by Davisson and Germer for electron energies up to 370 electron-volts. In each case, the agreement between the observed and the computed wave lengths was within the limits of error of the observations. Subsequently, with improved apparatus, they reported measurements in which the agreement between observed and computed values of wave length was better than 1 percent.³

The excellent quantitative agreement between such directly observed values of λ_e with values computed from de Broglie's formula is conclusive proof that the beam of electrons proceeding from filament to crystal does have wavelike characteristics as definite as those of

¹ The difference in the colatitude at which the spur is most strongly developed for the *A* and the *B* azimuth, respectively, is due to the fact that the displacement of each plane grating with respect to the one immediately above is different for the two azimuths, and hence the angle at which greatest reinforcement occurs is different in the two cases.

² Plane of incidence and reflection parallel to a side of the triangle.

³ *Nat. Acad. Sci., Proc.*, vol. 14, p. 317 (1928).

light or X-rays; and, furthermore, that the length of the electron waves is correctly given by de Broglie's formula, Eq. (167).

(b) *Refraction in a Crystal*.—When Davisson and Germer let the electrons fall *obliquely*, instead of perpendicularly, upon the nickel

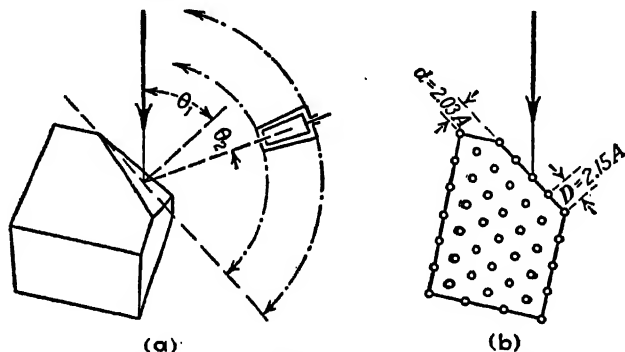


FIG. 83.—(a) Schematic representation of apparatus for studying reflection of a beam of electrons incident at angle θ_1 , on the (out) face of the nickel crystal. (b) Cross section of the crystal showing the "cooperating" planes of atoms.

crystal, they observed new effects which did not seem to agree entirely with the theory.¹ Their figures illustrating this case are reproduced in Fig. 83. With a fixed angle of incidence θ_1 , electrons were observed to be reflected at all angles θ_2 , as is illustrated in Fig. 84, which is also taken from their paper. For any given angle of incidence, however,

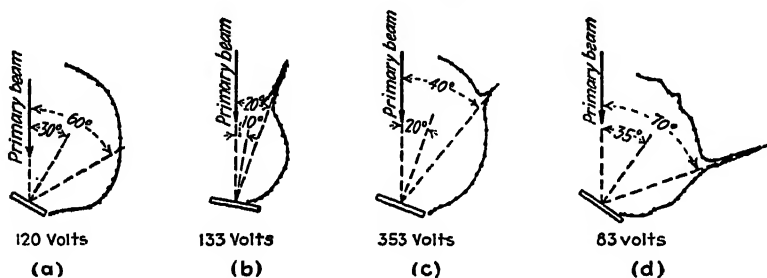


FIG. 84.—(a) Nonselective reflection of an electron beam. (b, c, d) Selective reflection at various angles of incidence.

at certain electron speeds, *selective regular reflection* occurred, as is shown by the "spurs" in Fig. 84(b), (c), and (d).

It was assumed that this selective reflection was due to constructive interference by beams reflected from successive layers of atoms in the crystal. The phenomenon is well known in the case of X-rays, which are very weakly reflected unless the conditions are met for such con-

¹ DAVISSON and GERMER, *Nat. Acad. Sci., Proc.*, vol. 14, pp. 317, 619 (1928).

structive interference. The condition for strong reflection will be deduced later for use in the more important case of X-rays; it is expressed to a high degree of approximation by Bragg's formula:

$$n\lambda = 2d \sin \theta, \quad (169)$$

where λ is the wave length, θ , the "glancing angle" or angle between the incident ray and the plane of reflection (complement of the angle of incidence), d the distance between layers [shown as 2.03 Å in Fig. 83(b)], and n a positive integer. (See Sec. 186.) If we may assume that the electron waves penetrate a short distance into the crystal, so as to be reflected from a number of layers of atoms, the same law should hold for the selective reflection of electrons. (In this case, the spacing in a given layer, shown in Fig. 83(b) as $D = 2.15$ Å, is of no consequence, since reflection from each layer is regular.)

When Davisson and Germer calculated electron wave lengths from their data by means of the Bragg formula, however, a discrepancy was found between them and the de Broglie wave lengths. Thus, for 83-volt electrons, selective reflection was observed at an angle of incidence of 35° [Fig. 84(d)]; putting $\theta = 90 - 35 = 55^\circ$, $d = 2.03$ Å, in Eq. (169), we find

$$\lambda_e = \frac{3.33}{n} \text{ Å.}$$

For 83-volt electrons, Eq. (167) gives a de Broglie wave length of

$$\lambda_e = 1.34 \text{ Å.}$$

It is impossible to find an integral n (1, 2, . . .) that will bring these two equations into agreement (if $n = 2$, $\lambda_e = 1.66$ Å; if $n = 3$, $\lambda_e = 1.11$ Å).

It was suggested by Eckart¹ and by Bethe² that this discrepancy would be explained if it were assumed that the electron beam suffers *refraction* as it enters the crystal. In the case of X-rays, the refractive index is so nearly unity that it need be taken into consideration only in very precise work (Sec. 199). In the latter case, Eq. (169) must be replaced by

$$n\lambda = 2d \sqrt{\mu^2 - \cos^2 \theta}, \quad (170)$$

μ being the refractive index for rays entering the crystal.

Equation (170) will presumably hold also for electron waves. Since we do not know μ , we cannot use it to secure another check of

¹ ECKART, *Nat. Acad. Sci., Proc.*, vol. 13, p. 460 (1927).

² BETHE, *Naturwiss.*, vol. 15, p. 787 (1927).

the de Broglie relation; but, if we assume the latter relation, we can use the formula to calculate μ . It is necessary, however, to make some assumption concerning the values of n . The right assumption to make might be discovered by endeavoring to secure consistency between values of μ obtained from many different reflections; but it is more satisfactory to be guided by physical considerations based upon the underlying cause of the refraction. In Sec. 114, we saw that electron waves are refracted upon entering a region of different electric potential; if the potential suddenly increases by an amount ΔV , so that the potential energy of the electron decreases by $e\Delta V$, e being the numerical electronic charge, the index of refraction is

$$\mu = \sqrt{1 + \frac{\Delta V}{V_0}}, \quad (171)$$

V_0 being the energy of the electrons before entering the region, expressed in equivalent potential units. Now photoelectric and thermionic phenomena indicate that matter *attracts* electrons; the potential energy of an electron decreases when it enters a metal, by an amount that was formerly identified with the work function but is now estimated at 10 to 20 electron-volts (Sec. 57). Hence μ should be greater than unity for electron waves entering a crystal; and ΔV as calculated from Eq. (171) should be of the order of 10 to 20 volts.

Using the above data, *i.e.*, $\lambda = 1.34$ Å, $d = 203$ Å, $\theta_g = 55^\circ$, if we take $n = 2$, we find, from (170), $\mu = 0.76$. This is less than unity and, for the reason just stated, must be rejected. Trying $n = 3$, we find $\mu = 1.144$; and (171) then gives, with $V_0 = 83$ volts, $\Delta V = 26$ volts, which is of the order to be expected.

A careful study of the selective reflections occurring for various voltages at the angle of incidence $\theta_1 = 10^\circ$ ($\theta_g = 80^\circ$) was made by

TABLE I.—INDEX OF REFRACTION OF NICKEL FOR ELECTRON BEAMS OF VARIOUS VOLTAGES. $\theta_g = 80^\circ$

V , volts	λ , angstroms	Order of diffraction, n (assumed)	μ [from Eq. (170)]	$\mu = \sqrt{1 + \frac{\Delta V}{V_0}}$ for $\Delta V = 18$ volts
64	1.52	3	1.14	1.132
130	1.07	4	1.07	1.068
215	.83	5	1.04	1.042
327	.67	6	1.02	1.027
450	.58	7	1.01	1.020
585	.50	8	1.008	1.015

Davisson and Germer. A summary of the data is shown in the first four columns of Table I. The values of the index of refraction computed by Eq. (171) for an assumed inner potential ΔV of 18 volts for the nickel crystal are shown in column 5 and are seen to be in reasonable agreement with the directly observed values of μ as given in column 4.

The phenomena of the refraction of electron waves are somewhat more complex, however, than might be expected from the simplicity

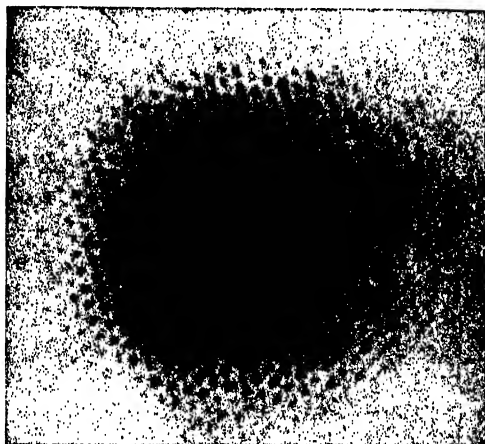


FIG. 85.—Diffraction of 68,000-volt electrons by a thin sheet of mica. [After Kikuchi, *Japan. Jour. Physics*, vol. 5 (1928).]

of Eq. (171) and the agreement of the data. For further information, the reader is referred to the literature.¹

(c) *Transmission through a Crystal*.—Electrons, if given sufficient energy, are known to pass readily through thin films of matter such as metal foil or thin mica. If an electron beam has wave properties, then in its passage through matter we should observe some or all of the phenomena characteristic of the passage of X-rays through matter.

Soon after the discovery of electron waves, Kikuchi² succeeded in obtaining electron diffraction patterns by passing a pencil of electrons through a thin mica crystal. Figure 85 shows the pattern obtained by passing 68,000-volt electrons through mica of the order of 10^{-5} cm. thick. These observations are the exact analog of the

¹ See, for example, DAVISSON and GERMER, *Nat. Acad. Sci., Proc.*, vol. 14, p. 619 (1928); DAVISSON: *Bell System Techn. J.*, vol. 11, p. 546 (1932).

² *Japan. Jour. Phys.*, vol. 5, p. 83 (1928).

first experiments on the diffraction of X-rays by Friedrich, Knipping, and Laue [Sec. 186(b)].

Studies of metal foils with X-rays show that with appropriate treatment such foils are composed of polycrystalline material with a random orientation of crystal axes. Crystallographically, such foils are similar to crystalline powders. If in the Friedrich, Knipping, and Laue experiment with X-rays (Fig. 150) a powder or a foil is substituted for the single crystal, the observed diffraction pattern will consist of a series of concentric rings instead of spots as in the Laue

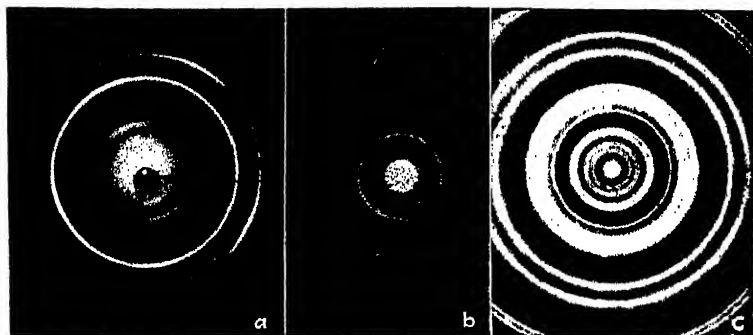


FIG. 86.—(a) Diffraction pattern produced by passing X-rays through thin gold foil (*Kindness of Dr. I. Fankuchen*). (b) Diffraction pattern produced by passing a pencil of 30,000-volt electrons through gold foil (*from Thomson: "Wave Mechanics of Free Electrons," Fig. 19b*). (c) Same for 210,000-volt electrons (*from Rupp: Ann. d. Physik, vol. 10, p. 927 (1931)*).

pattern. Such a pattern, produced by passing a pencil of X-rays through gold, is shown in Fig. 86(a). In Fig. 86(b) is shown the diffraction pattern obtained by passing 30,000-volt electrons through a film of gold; whereas Fig. 86(c) shows the pattern for 210,000-volt electrons, obtained by E. Rupp. Qualitatively, the similarity between the effect of X-rays and of electrons is so striking as to prove conclusively the wave nature of electrons. From the diameter of the rings in Fig. 86(b) and the dimensions of the apparatus, it is possible, knowing the length of the side of the unit cube of gold from X-ray data, to measure the wave length of the electron beam producing the pattern. This can be compared with the value of the wave length computed from the relativistic formula (page 272, footnote) appropriate to high-velocity electrons. In this way, Rupp¹ compared

¹ The validity of Rupp's results is not beyond possible doubt. Cf. RUPP, *Zeits. f. Physik*, vol. 95, p. 801 (1935); RAMSAUER, *Zeits. f. Physik*, vol. 96, p. 278 (1935).

the "observed" with the "computed" values of electron waves in the voltage range 100 to 250 kilovolts. His results are summarized in Table II.

(d) *Diffraction by a Ruled Grating.*—In concluding our citation of evidence to show that electron beams have wave characteristics,

TABLE II.—COMPARISON OF DIRECTLY OBSERVED AND COMPUTED VALUES OF ELECTRON WAVE LENGTHS FOR HIGH-VOLTAGE ELECTRONS*

Maximum energy of electron beam in electron-volts $\times 10^{-4}$	Wave length, cm. $\times 10^{11}$	
	Computed	Observed
11.98	32.5	32.52
21.0	23.6	23.72
23.1	22.4	22.51
24.3	21.7	21.42
25.3	21.2	21.06

* RUPP: *Ann. d. Physik*, vol. 10, p. 927 (1931).

mention should be made of the experiments of Rupp¹ with a ruled grating. Allowing 150-volt electrons to fall at a small glancing angle on a line grating of 1,300 lines per centimeter ruled on speculum metal, Rupp obtained on a photographic plate, in addition to the direct-reflected beam, three orders of diffraction, from which the wave length was determined to be 1.00 angstrom with an error of 2 or 3 percent—in excellent agreement with the expected value of 0.999 Å for 150-volt electrons.

118. Diffraction of Molecule Waves.—Even *molecules* should exhibit wave properties under suitable conditions, according to the new theory. This, too, has been verified by experiment. A molecular beam is easily formed by allowing molecules of a gas to stream out of an enclosure through a small hole or slit into an evacuated chamber. Often the enclosure is heated, in order to vaporize the substance to be studied; it is then called an "oven." Two difficulties have to be overcome, however, which do not arise in working with electrons. The molecules issue with a maxwellian distribution of velocities, whereas for diffraction experiments a beam of uniform velocity is desirable, corresponding to monochromatic waves. Then, too, neutral molecules are very much harder to detect than charged particles.

¹ *Zeits. f. Physik*, vol. 52, p. 8 (1929).

Especially interesting is an experiment¹ performed by Estermann, Frisch, and Stern. They managed to select a beam of helium atoms having fairly uniform velocities by passing the beam through narrow slits in two parallel circular disks placed 3 cm. apart and rotated about their common axis. The slits in the disks were adjusted so as to be opposite each other. A molecule, after passing through a slit in the first disk, would arrive at the second disk too late to pass through the corresponding slit in that disk; but, if its velocity were just right, it would be in time to pass through the *next following slit*. Molecules moving faster or more slowly would arrive too soon or too late and would be stopped by the disk. After leaving the second disk, the beam fell upon the surface of a lithium fluoride crystal, by which it was reflected or diffracted. To measure the intensity of the beam diffracted in a given direction, the helium atoms were allowed to pass through a small hole into a chamber where they accumulated until a certain pressure of helium was reached (of the order of 10^{-6} cm. Hg). This pressure was measured by the cooling effect of the helium on an electrically heated metal strip, owing to conduction of heat through the helium; the electrical resistance of the strip was measured with a Wheatstone bridge as an indication of its temperature.

A strong diffracted beam of helium molecules or atoms was observed in addition to the regularly reflected beam. In the most precise measurement, the maximum in the diffracted beam was found at 19.45° , corresponding to a wave length of $\lambda = 0.600 \times 10^{-8}$ cm. as calculated from the usual formula for diffraction by a crystal. From the dimensions and rate of rotation of the disk, the velocity of the helium atoms was calculated to be 1.635×10^5 cm. sec.⁻¹; the corresponding de Broglie wave length is, by Eq. (166a,b),

$$\lambda = \frac{h}{mv} = \frac{hN_0}{Mv} = \frac{6.61 \times 10^{-27} \times 6.02 \times 10^{23}}{4.00 \times 1.635 \times 10^5} = 0.608 \times 10^{-8} \text{ cm.}$$

(h = Planck's constant, N_0 = Avogadro's number, M = molecular weight.) The agreement is within 1.5 percent. This result is particularly interesting, not only because it refers to an atom, but also because the velocity was measured in simple mechanical fashion instead of in some indirect manner.

119. Schrödinger's Wave Equation.—In Secs. 112 and 113, we have described the line of thought which led de Broglie to the conception of matter waves. In his first paper, he managed to apply these ideas to the motion of an electron in the field of a nucleus and actually arrived at Bohr's formula for the energies of the quantum

¹ ESTERMANN, FRISCH and STERN, *Zeits. f. Physik*, vol. 73, p. 348 (1931).

states. In order to obtain discrete states, he replaced Bohr's quantum condition by the requirement that the orbit must contain an integral number of wave lengths of the waves, so that "resonance" may occur. This is not a very satisfactory statement of the condition, however, because in such cases the wave length is so large relative to the scale of the motion that diffraction effects would be enormous. Schrödinger, refining de Broglie's idea, suggested that each quantum state represents a system of *standing waves*, or a *normal mode* of harmonic vibration, somewhat like the vibration of a violin string when sounding its fundamental, or one of its overtones.

In order to find out what the normal modes of oscillation would be for the matter waves accompanying an electron, we require a more complete statement of the law governing their motion. Now the usual basis of the mathematical theory of wave motion is furnished by a *differential equation*. Thus, for a violin string vibrating in a fixed plane, we have the equation

$$\frac{\partial^2 \eta}{\partial t^2} = g^2 \frac{\partial^2 \eta}{\partial x^2},$$

in which η is the displacement of any point of the string at a time t , x is the coordinate of this point measured from one end of the string, and g is a constant depending on the tension in the string and its mass per unit length. In addition to this differential equation, we have also the *boundary condition* that $\eta = 0$ at all times at each end of the string. To cite a second example, in the case of sound waves moving in three dimensions, the differential equation for the pressure p in terms of Cartesian coordinates x, y, z and the time t is

$$\frac{\partial^2 p}{\partial t^2} = f^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right),$$

f denoting another constant. These differential equations are derived from the laws of mechanics; they are obtained by applying Newton's laws of motion to the particular medium under discussion.

We cannot follow an analogous method, however, in arriving at the differential equation for the matter waves; for we know nothing of any medium to convey them. The best we can do is to try to guess a mathematical wave equation that will be in harmony with those properties which we have found the waves to have, and then to test our guess further by comparing with observation other results deduced from this wave equation. Limitations of space and of the mathematical tools required for this book make it impossible for us to

attempt here any extensive treatment of wave mechanics.¹ It may be worth while, however, to see how the properties that we have assigned to matter waves lead naturally to Schrödinger's differential equation for them and then to describe in general terms the application of the equation to a few important problems.

For plane harmonic matter waves of frequency ν and wave length λ moving toward $+x$, we might expect to be able to write

$$\Psi = A \sin 2\pi \left(\nu t - \frac{x}{\lambda} \right),$$

Ψ being the quantity, whatever it is, that oscillates in the waves. Since, according to Eqs. (165) and (166a),

$$\nu = \frac{W}{h}, \quad \lambda = \frac{h}{p},$$

where W and p are, respectively, the energy and the momentum of the associated particle, we can also write the last equation in the form

$$\Psi = A \sin \frac{2\pi}{h} (Wt - px), \quad (172)$$

in which the constants are connected by the relation,

$$W = \frac{p^2}{2m} + V; \quad (173)$$

here V is the potential energy of the particle, and

$$\frac{p^2}{2m} = \frac{(mv)^2}{2m} = \frac{1}{2}mv^2,$$

its kinetic energy. Whatever the differential equation for Ψ may be, it should be such that Ψ as given in Eq. (172) is a solution of it regardless of the value of W or p . Let us endeavor, therefore, to find by trial a differential equation such that, when the expression just given for Ψ is substituted in it, W and p cancel out in consequence of the relationship expressed in Eq. (173).

Let us write down a few derivatives of Ψ as obtained from Eq. (172):

$$\begin{aligned} \frac{\partial \Psi}{\partial t} &= \frac{2\pi}{h} W A \cos \frac{2\pi}{h} (Wt - px) & \frac{\partial \Psi}{\partial x} &= -\frac{2\pi}{h} p A \cos \frac{2\pi}{h} (Wt - px) \\ \frac{\partial^2 \Psi}{\partial t^2} &= -\frac{4\pi^2}{h^2} W^2 A \sin \frac{2\pi}{h} (Wt - px) & \frac{\partial^2 \Psi}{\partial x^2} &= -\frac{4\pi^2}{h^2} p^2 A \sin \frac{2\pi}{h} (Wt - px). \end{aligned}$$

¹ For a simple introduction, see S. Dushman, "Elements of Quantum Mechanics," 1938.

The only sort of combination of these derivatives that would enable us to use Eq. (173) seems to be one in which the *first* derivative with respect to t is combined with the *second* derivative with respect to x . We can get a term in V by using Ψ itself, undifferentiated, multiplied by V . We are thus led to try as a wave equation

$$\frac{\partial \Psi}{\partial t} = a \frac{\partial^2 \Psi}{\partial x^2} + b V \Psi, \quad (174)$$

where the coefficients a and b remain to be determined.

If we substitute Ψ from (172) in this equation, however, we encounter a difficulty, since $\partial \Psi / \partial t$ contains a cosine, whereas $\partial^2 \Psi / \partial x^2$ and Ψ itself both contain a sine. However, we could equally well have *started* with a cosine instead of a sine; or we could use a *combination of them*, such as

$$\Psi = A \sin \frac{2\pi}{h}(Wt - px) + B \cos \frac{2\pi}{h}(Wt - px),$$

where A and B are independent constants. The student may already be familiar with the use of such combinations as the general solution of a differential equation of the second order. If we substitute this latter expression for Ψ in (174) and then equate coefficients of the sine terms and the cosine terms separately on both sides of the equation, so as to have it satisfied for all values of x and t , we obtain the equations:

$$\begin{aligned} -\frac{2\pi}{h} WB &= -a \frac{4\pi^2}{h^2} p^2 A + bVA, \\ \frac{2\pi}{h} WA &= -a \frac{4\pi^2}{h^2} p^2 B + bVB, \end{aligned}$$

or

$$W = a \frac{2\pi}{h} p^2 \frac{A}{B} - \frac{bh}{2\pi} V \frac{A}{B} = -a \frac{2\pi}{h} p^2 \frac{B}{A} + \frac{bh}{2\pi} V \frac{B}{A}.$$

These equations will agree with (173), provided

$$\begin{aligned} a \frac{2\pi}{h} \frac{A}{B} &= \frac{1}{2m}, & -\frac{bh}{2\pi} \frac{A}{B} &= 1, \\ -a \frac{2\pi}{h} \frac{B}{A} &= \frac{1}{2m}, & \frac{bh}{2\pi} \frac{B}{A} &= 1. \end{aligned}$$

Eliminating either a or b from these equations, we obtain

$$\frac{A^2}{B^2} = -1, \quad \therefore \frac{A}{B} = \pm i; \quad i = \sqrt{-1}.$$

To agree with common practice, let us take

$$\frac{A}{B} = -i.$$

Then we obtain (since $1/i = -i$):

$$-ia \frac{2\pi}{h} = \frac{1}{2m}, \quad a = \frac{i\hbar}{4\pi m}; \quad i \frac{b\hbar}{2\pi} = 1, \quad b = \frac{2\pi}{i\hbar}.$$

Thus (174) becomes

$$\frac{\partial \Psi}{\partial t} = \frac{i\hbar}{4\pi m} \frac{\partial^2 \Psi}{\partial x^2} + \frac{2\pi}{i\hbar} V\Psi$$

or, after multiplying through by $-h/(2\pi i)$,

$$-\frac{h}{2\pi i} \frac{\partial \Psi}{\partial t} = -\frac{h^2}{8\pi^2 m} \frac{\partial^2 \Psi}{\partial x^2} + V\Psi.$$

In three dimensions, this will read

$$-\frac{h}{2\pi i} \frac{\partial \Psi}{\partial t} = -\frac{h^2}{8\pi^2 m} \left(\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} \right) + V\Psi$$

or, in a condensed notation often used,

$$-\frac{h}{2\pi i} \frac{\partial \Psi}{\partial t} = -\frac{h^2}{8\pi^2 m} \nabla^2 \Psi + V\Psi, \quad (175)$$

∇^2 (or Δ) standing for the "Laplacian operator" or the sum of the three space derivatives.

Equation (175) is Schrödinger's famous *wave equation containing the time*. It appears to be valid for the matter waves associated with 1 electron as long as relativistic effects can be neglected.

Schrödinger's equation is simple in form, containing in its coefficients only the potential-energy function for the particle, $V(x, y, z)$, and the universal constant \hbar . It is remarkable among the differential equations of mathematical physics, however, in that it contains $i = \sqrt{-1}$. But for this factor it would resemble the equation for the flow of heat in a solid body, which likewise contains the first derivative with respect to the time. The motion of matter waves combines, in fact, some of the features of heat flow in a solid with other features resembling the propagation of mechanical disturbances, such as sound waves.

The *solutions* of the equation are likewise peculiar. As we have seen, no solution can have the form of a simple sine or cosine function, such as can be used to represent a train of sound waves in the air

[cf. Eq. (172)]. The combination of sine and cosine that we found ourselves compelled to use is equivalent to a complex exponential, since $e^{\pm ix} = \cos x \pm i \sin x$.

In case V is a constant, it is easily seen, in fact, by substitution, that a solution of (175) is

$$\Psi = C e^{-\frac{2\pi i}{h}(Wt - px)} \quad (176)$$

provided W and p are constants having such values that

$$W = \frac{p^2}{2m} + V.$$

Here C is an arbitrary constant representing the wave amplitude, and $i = \sqrt{-1}$. If $W > 0$ and $p > 0$, this latter form of Ψ represents plane waves of Ψ traveling toward $+x$; for, at a point moving with velocity W/p , the value of Ψ remains constant.¹ W and p represent energy and momentum of the associated particle, respectively; the frequency ν and the wave length λ of the waves are easily seen to be $\nu = W/h$, $\lambda = h/p$, in accordance with conclusions reached previously. If $p < 0$ but $W > 0$, both particle and waves are moving toward $-x$.

If $W < 0$, as may happen if $V < 0$, the waves and the particle travel in opposite directions. We can always make this so, if we wish, by adjusting the arbitrary additive constant that occurs in V , as in any potential energy. The possibility of reversing the direction of motion of matter waves in this manner is a clear indication that they cannot be ordinary real waves.

120. Physical Significance of Ψ .—This may be an appropriate point at which to discuss further the physical nature of matter waves. What is the physical significance of the mysterious Ψ that appears in the mathematical theory?

An answer sometimes given to this question is that Ψ is merely an auxiliary mathematical quantity that is introduced to facilitate computations relative to the results of experiment. For example, in the experiments described above on the diffraction of electrons by a crystal, the experimenter sets up an electron gun that fires electrons of a certain energy at a crystal. A detector placed at a certain angle gives indications which are taken to mean that electrons are being received. In order to develop a quantitative theory for such observations, we assume that a beam of electron waves of frequency $\nu = W/h$ falls on the crystal, and we calculate the intensity of the waves scat-

¹ If a point moves with velocity W/p , in time dt it moves a distance $dx = Wdt/p$; hence, $d(Wt - px) = Wdt - p dx = 0$, and so $d\Psi = 0$.

tered in the direction of the detector. From this mathematical result we infer, following certain rules, what the indication of the detector should be. In order to make such calculations, it is not really necessary to attach any physical significance at all to the mathematical symbol Ψ .

There is much to be said for such a view. After all, observational results are the primary material of physics; the purpose of theories is to correlate these results, to group them into those regularities of experience that we call laws, and to predict the results of new experiments. Even the motion of the electron as a particle is only an auxiliary concept, introduced for convenience in describing and interpreting observational results; the evidence for the particle nature of the electron is in reality very indirect. If we carefully examine the history of the concept of the electron as a particle, we shall realize that this concept owes its origin to the well-established beliefs in the corpuscular nature of matter. The electron was regarded as one of the building blocks of matter and was, therefore, visualized as a particle.

Such auxiliary concepts as particles constitute very convenient aids to thinking, and most physicists find it advantageous to make use of them. Whereas some experimental results are conveniently understood in terms of the electrons as particles, however, others can be understood only in terms of the interference of waves. It is worth while, therefore, to go as far as we can toward assigning some physical significance to these waves, or to the quantity Ψ in terms of which the mathematical theory of the waves is expressed. This can be done, according to circumstances, in two rather different ways.

In dealing with an experiment on scattering such as that just described, a beam of waves of indefinite total length is commonly assumed, represented by a Ψ like that in Eq. (176). The *square of the absolute value* of Ψ , $|\Psi|^2$, is then taken to be proportional (for given particle energy) to the number of particles in the beam that cross unit area per second, the unit area being taken perpendicular to the direction motion.¹ Thus, in an experiment on the diffraction of electrons, let n_0 electrons cross unit area per second in the incident beam and let Ψ_0 denote the mathematical expression assumed for Ψ in the waves representing this beam. Then, if calculation by means of the wave equation, Eq. (175), gives Ψ_d as the value of Ψ in a certain direction of diffraction, the theoretical value for the intensity of the diffracted beam in that direction, defined as the number of diffracted electrons

¹ Ψ is a complex number; by the absolute value of any complex number, $x + iy$ (where x and y are real), is meant $(x^2 + y^2)^{1/2}$, written $|x + iy|$.

crossing unit area per second, is

$$n_d = \frac{|\Psi_d|^2}{|\Psi_0|^2} n_0$$

[provided both the incident and the diffracted beams move in regions where V in (175) has the same value].

This procedure corresponds exactly to the method of handling similar problems in optics. The intensity of a beam of light, which might be defined as the number of photons crossing unit area per second, is proportional (in a given medium) to E^2 , the square of the electric vector (Sec. 34). The principal difference here is that, Ψ being a complex number, we must use $|\Psi|^2$ instead of Ψ^2 .

There are other cases, however, as in the photoelectric effect, in which one wishes to follow the flight of a single electron or other particle. Then Ψ is taken to refer to a single particle. Usually values of Ψ appreciably different from zero occur only within some finite region. A solution of the wave equation of this latter type is sometimes called a *wave packet*, especially if the region in which it has nonzero values changes position in space as time goes on. In such a case it is natural to ask, where is the particle in relation to the wave packet? The accepted answer may be stated in terms of probabilities. The *position of the particle* is believed *not to be defined any more closely than is indicated by the values of Ψ* . At any given instant, if a suitable observation were made, the particle might be found at any point where Ψ is different from zero; the *probability* of finding it at a given point is proportional to the value of $|\Psi|^2$ at that point. More exactly, the *probability of finding the particle within any element of volume $dx dy dz$* is

$$|\Psi|^2 dx dy dz. \quad (177)$$

When interpreted in this way, Ψ is sometimes called a *probability amplitude* for position of the particle.¹

The interpretation just stated imposes upon Ψ a certain mathematical requirement. For the *total probability* of finding the particle *somewhere* is, of course, unity. Hence Ψ must satisfy the condition that

$$\iiint |\Psi|^2 dx dy dz = 1, \quad (178)$$

the triple integral extending over all possible values of x , y , and z . A Ψ satisfying this requirement is said to be *normalized* (to unity). Any solution of Eq. (175) can be normalized; for, because the wave

¹ The word "amplitude" is not used here in quite the same sense as in elementary physics.

equation is linear and homogeneous in Ψ , any solution of this equation can be multiplied by an arbitrary constant without ceasing to be a solution, and the arbitrary constant can then be adjusted so as to make Eq. (178) true.

There are also interesting relations between a wave packet and the *velocity* or *momentum* of the particle. A detailed study, for which we have not space, brings out the following features.

Suppose the particle is in free space (*i.e.*, $V = 0$). Then a possible solution of the wave equation is given by Eq. (176) or

$$\Psi = C e^{-\frac{2\pi i}{h}(Wt - px)}, \quad (179)$$

where C and p are constants and $W = p^2/m$. This solution, representing an infinite train of waves, is considered to correspond to a particle moving with energy W , momentum p , and velocity $v = p/m$. It cannot be treated as a wave packet, however, for it makes

$$\iiint |\Psi|^2 dx dy dz = \infty$$

(unless $C = 0$). We may conclude that *it is not possible for a particle to have a perfectly definite momentum or velocity*. We can, however, have as close an approach as we please to a definite momentum, for Ψ may be indistinguishable from (179) over as large a region as desired, sinking to zero outside the boundary of this region; the constant C can then be chosen small enough so that $\iiint |\Psi|^2 dx dy dz = 1$. A wave packet formed in this way will retain its form for a long time, traveling along with a velocity roughly equal to the group velocity of waves of frequency $\nu = W/h$ or with a velocity equal to p/m .

In general, any wave packet can be expanded in terms of wave trains like (179), just as any patch of light waves can be resolved into monochromatic trains. Essentially, this amounts to representing Ψ by a Fourier integral. When this has been done in a suitable way, the coefficients of the various wave trains in the expansion constitute a probability amplitude for momentum; *i.e.*, the square of the absolute value of any coefficient gives the probability that a suitable observation would reveal the particle as moving in the direction, and with the momentum or velocity, that is associated with the corresponding wave train. Thus, in general, according to wave mechanics, a particle does not have either a sharply defined position or a sharply defined momentum or velocity.

121. The Indeterminacy Principle.—The indefiniteness that we have just found to exist in the values of certain mechanical magnitudes associated with a particle, such as its position or momentum, is a

fundamental feature of wave mechanics. The indefiniteness as to position can be minimized, to be sure, by making the wave packet very small (Ψ practically zero except within a very small region); but in that case it can be shown that the packet will spread rapidly. Consequently, if we were to observe the position of the particle a little later and then calculate its velocity by dividing the distance covered by the time taken, any one of a wide variety of results might be obtained. Thus, a small packet means a large indefiniteness in momentum and velocity. On the other hand, if we give to Ψ a form like that in Eq. (179) over a large region, in order to fix the velocity and momentum of the particle within narrow limits, there is a large indefiniteness in the position. In general, it can be shown that, if Δx denotes the effective range in the possible values that might be found by observation for the coordinate x of the particle, and if Δp indicates the similar range for momentum, then very roughly

$$\Delta p \Delta x \geq h, \quad (180)$$

where h is Planck's constant. This principle was first enunciated by Heisenberg in 1927,¹ who called it, in German, the principle of "Unbestimmtheit." This term has been variously translated as "indeterminacy," "indefiniteness," "uncertainty."

The principle can be said to have its basis in the *wave properties* of matter. It even has an analog in the field of optics (but not in particle mechanics). A single sinusoidal wave of light of wave length λ represents a certain amount of energy that is closely localized in space, but it does not constitute monochromatic light; for, upon passing through a spectroscopic slit, it will spread widely in the spectrum. To have an approach to monochromatic light, we must have a train of many waves; and this means a corresponding dispersal of the energy in space.

The conclusion that a wave packet cannot represent a particle as having at the same time a definite position and a definite momentum might seem to be in conflict with the fact that both position and momentum are capable of precise measurement. In ordinary physics, we have no difficulty in determining both of these quantities simultaneously; *e.g.*, from two snapshots of a rifle bullet, its position and velocity at a given instant can both be calculated. Heisenberg pointed out, however, that this could be done only because, on the scale of observation used in ordinary physical measurement, the indeterminacy required by Eq. (180) is so minute as to be lost in the experimental errors. It is quite otherwise for an electron, or a molecule.

¹ HEISENBERG, *Zeits. f. Physik*, vol. 43, p. 172 (1927).

Consider, for example, how an electron might be located with *atomic* precision. We might use a microscope; but then we should have to use light of extremely short wave length in order to secure sufficient resolving power. To distinguish positions 10^{-9} cm. apart, for example, we should have to use γ -rays. Under these circumstances, the effect of the light on the electron cannot be neglected. If we are to "see" the electron, at least 1 photon must bounce off it and enter the microscope. In rebounding from the electron, however, this photon will give it a strong Compton kick (see Sec. 198). Thus at the instant at which we locate the electron, its momentum undergoes a discontinuous change. Furthermore, there is an indefiniteness about the magnitude of this change, for it will vary according to the direction in which the scattered photon leaves the scene of action. We cannot limit closely the range of possible directions for the scattered photons that enter the microscope, by stopping down the aperture, without a serious loss of resolving power. A quantitative analysis of such observations leads us again to Eq. (180).

It appears, then, that we cannot *at the same time* assign to an electron or other small particle a definite position and a definite momentum (or energy). At least we cannot do this and mean anything by the statement in terms of physical observation; and, since 1900, physicists have become increasingly clear about the principle that only those magnitudes which can be observed, directly or indirectly, have physical significance. Thus, our classical notion of a particle as something that can move along a sharply defined path, having at each instant a definite position and velocity, is not applicable to electrons or protons or atoms or molecules. These small bits of matter may be said to have some particle properties, but they also possess certain wave properties, so that, in the classical sense of the words, they are neither true particles nor true waves. Darwin proposed to call them "wavicles."

After this discussion of the physical significance of Ψ , we shall now return to a consideration of the mathematical theory. Whatever changes may occur in the future in the physical ideas that are considered to underlie wave mechanics, it seems certain that the mathematical theory is here to stay.

122. Stationary States.—So far we have discussed the motion of free particles only. Suppose now an electron or other particle is in a region where, because of a force field, it has a potential energy V depending on its position. To treat such cases, as has already been indicated, Schrödinger suggested seeking solutions of the wave equation which represent standing waves.

The distinguishing characteristic of standing waves is that the phase of the vibration in such waves is everywhere the same, whereas in running waves there is a progression of phase along the wave train. In a mathematical expression for Ψ representing standing waves, therefore, the time must occur in a separate factor. Let us seek a solution of Eq. (175) that is of the form

$$\Psi(x, y, z, t) = e^{-\frac{2\pi i}{h} W t} \psi(x, y, z). \quad (181)$$

Here W is a constant, $i = \sqrt{-1}$, h is Planck's constant, t is the time, and x , y , and z are Cartesian coordinates for the particle; Ψ , as indicated, is a function of the four variables, x , y , z , t , whereas ψ is a function of only x , y , and z . The frequency of oscillation of such a Ψ is $\nu = W/h$. For, if t is increased by h/W , the exponent of e is thereby increased by $-2\pi i$, which multiplies Ψ by

$$e^{-2\pi i} = \cos 2\pi - i \sin 2\pi = 1$$

and so leaves Ψ unchanged; hence h/W is the period of oscillation of Ψ , or W/h is the frequency. Since the frequency of harmonic matter waves always represents the energy divided by h (Secs. 112, 113, 116), it follows that W represents the energy of the particle. It is easily shown that Eq. (181) defines the only possible form of Ψ in which t occurs in a separate factor.

Differentiating, we find from (181) that

$$\frac{\partial \Psi}{\partial t} = -\frac{2\pi i}{h} W e^{-\frac{2\pi i}{h} W t} \psi, \quad \frac{\partial^2 \Psi}{\partial x^2} = e^{-\frac{2\pi i}{h} W t} \frac{\partial^2 \psi}{\partial x^2}, \text{ etc.}$$

Thus the operator ∇^2 affects only the factor ψ . Hence, if we insert the above expression for Ψ in Eq. (175), we obtain

$$W e^{-\frac{2\pi i}{h} W t} \psi = -\frac{h^2}{8\pi^2 m} e^{-\frac{2\pi i}{h} W t} \nabla^2 \psi + V e^{-\frac{2\pi i}{h} W t} \psi.$$

Canceling $e^{-\frac{2\pi i}{h} W t}$, we can write this result as

$$-\frac{h^2}{8\pi^2 m} \nabla^2 \psi + V \psi = W \psi$$

or

$$\nabla^2 \psi + \frac{8\pi^2 m}{h^2} (W - V) \psi = 0. \quad (182)$$

The differential equation thus obtained for the amplitude factor $\psi(x, y, z)$ is known as Schrödinger's *amplitude equation*, or, more commonly, just as *Schrödinger's wave equation*.

Not every solution of Eq. (182) can be used in the manner indicated, however. As explained in Sec. 120, we must be able to make $\iiint |\Psi|^2 dx dy dz$ finite. Now, if Ψ has the form shown in (181), its absolute value is

$$|\Psi| = |e^{-\frac{2\pi i}{h} W}| |\psi| = |\psi|$$

since $|e^{-i\theta}| = 1$.* Thus only those solutions of (182) can be used to represent standing waves for which

$$\iiint |\psi|^2 dx dy dz$$

is finite. (The statement usually made is that ψ itself must remain finite even at infinity.) In general, this condition can be met only for certain values of W , which are known as *characteristic values* (or *eigenvalues*). These values of W are the allowed values of the energy corresponding to the stationary states, or quantum states, of the particle in the given force field. Corresponding to each allowed value of W there are one or more different functions ψ , obtained as solutions of Eq. (182); these are known as *characteristic functions* or *wave functions* (or *eigenfunctions*).

The *probability* of finding the particle in a given element of space $d\tau$, when it is in such a stationary or quantum state, is obviously given either by $|\Psi|^2 d\tau$ or by $|\psi|^2 d\tau$, since, as we have seen, $|\Psi| = |\psi|$.

The problem of determining the allowed energy levels for a particle in a given force field thus reduces to the mathematical problem of finding the allowed solutions of a wave equation. Limitations of space and of the mathematical tools required for this book compel us to leave it to the reader to familiarize himself, by study elsewhere,¹ with the applications of the theory to the various problems of atomic and molecular physics. We have space only for a few descriptive remarks concerning the two cases that have already been treated in previous chapters in terms of the old quantum theory.

123. The Harmonic Oscillator.—Let x be the coordinate of a mass m moving in one dimension with a potential energy given by $V = \frac{1}{2}\beta x^2$, β denoting a constant. For such a particle, the (amplitude) wave equation (182) becomes simply

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2m}{h^2} \left(W - \frac{1}{2}\beta x^2 \right) \psi = 0. \quad (183)$$

* $|e^{-i\theta}| = |\cos \theta - i \sin \theta| = (\cos^2 \theta + \sin^2 \theta)^{1/2} = 1$.

¹ E.g., in DUSHMAN, S., "Elements of Quantum Mechanics," 1938; PAULING, L., and WILSON, "Introduction to Quantum Mechanics," 1935; ROJANSKY, V., "Introductory Quantum Mechanics," 1938.

A mathematical study of this equation shows that the characteristic values of W can be written¹

$$W_n = (n + \frac{1}{2})h\nu_c, \quad n = 0, 1, 2, \dots; \quad \nu_c = \frac{1}{2\pi}\sqrt{\frac{\beta}{m}} \quad (184a,b,c)$$

The constant ν_c is the frequency with which the mass m would vibrate according to classical theory; but it has no special physical significance here.

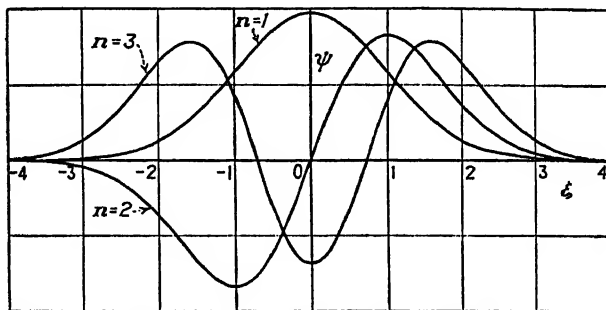


FIG. 87.—Characteristic functions for $n = 1, 2$, and 3 for a harmonic oscillator.

The normalized wave functions for the first three stationary states, corresponding to $n = 0, 1$, or 2 , are found to be as follows, stated in terms of

$$\xi = 2\pi x \sqrt{\frac{m\nu_c}{h}}, \quad C_0 = \left(2\sqrt{\frac{\pi m\nu_c}{h}}\right)^{1/2}.$$

For $n = 0$:

$$\psi_0 = C_0 e^{-\xi^2/2}; \quad n = 1: \quad \psi_1 = \sqrt{2} C_0 \xi e^{-\xi^2/2};$$

$n = 2$:

$$\psi_2 = \frac{C_0}{\sqrt{2}} (2\xi^2 - 1) e^{-\xi^2/2}.$$

The reader can readily verify that each of these expressions for ψ is a solution of Eq. (183), if in the latter equation the corresponding value of W is inserted from (184a), and that they are normalized so that $\int_{-\infty}^{\infty} \psi^2 dx = 1$. (Hint: change the variable from x to ξ even in the differential equation! Also, remember that $\int_{-\infty}^{\infty} e^{-\xi^2} d\xi = \sqrt{\pi}$.) A graph of these three wave functions is shown in Fig. 87.

124. The One-electron Atom. (a) *Energy Levels and Wave Functions.*—An electron of numerical charge e , in the field of a fixed nucleus

¹ *Ibid.*

of charge Ze , has a potential energy

$$V = -\frac{Ze^2}{r}$$

at a distance r from the nucleus. Hence the wave equation (182) for an electron in such a field reads

$$\nabla^2\psi + \frac{8\pi^2m}{h^2}\left(W + \frac{Ze^2}{r}\right)\psi = 0,$$

m being the electronic mass (more accurately, the reduced mass—Sec. 103). This equation is more complicated to handle than that for the harmonic oscillator, because it involves three variables (x, y, z). Only a few mathematical details will be given here. The variables can be “separated” in terms of polar coordinates, r, θ, ϕ ; solutions can then be found of the form

$$\psi_{nl\lambda} = C_{nl} e^{i\lambda\phi} P_l^\lambda(\cos\theta) R_{nl}(r). \quad (185)$$

Here $R_{nl}(r)$ is a certain function of r , P_l^λ denotes a certain function of $\cos\theta$, and C_{nl} is a normalizing factor. [Sometimes $R(r)/r$ is written in place of $R(r)$.] The symbols n, l , and λ denote three integers which occur in the customary expressions for the functions and also serve as quantum numbers or labels for them; these integers may have any values such that

$$|\lambda| \leq l < n.$$

Thus the possible values of n are $n = 1, 2, \dots$; for each n , $l = 0, 1, \dots, n-1$; for each l , $\lambda = -l, -l+1, -l+2, \dots, l-1, l$. (More commonly the letter m , or m_l , is used in place of λ . The notation used here has certain advantages which will appear in the study of many-electron spectra.) The corresponding values of the energy turn out to be the same as those found by Bohr (Sec. 103):

$$W = -\frac{2\pi^2me^4Z^2}{h^2n^2}. \quad (186)$$

This problem involves an important feature frequently met with, called *degeneracy* of the wave functions or quantum states. For any value of n above 1, there are several independent wave functions for each allowed value of the energy. For given l , λ ranges over $2l+1$ different values from $-l$ to l ; and, for given n , l ranges from 0 to $n-1$. The total number of functions or independent quantum states for given n is easily seen to be n^2 .

(b) *Relations with Angular Momentum.*—The different functions corresponding to the same energy are distinguished from one another by different values of the *angular momentum* of the electron about the nucleus, a feature that assists us in visualizing the quantum states. It can be shown that ψ_{nl} represents a state in which the angular momentum (or moment of momentum) about the axis of polar coordinates has the definite value

$$\lambda \frac{h}{2\pi}$$

For given l , this angular momentum can range by integral steps from a maximum of l units of $h/2\pi$ in one direction through zero to l units in the opposite direction, a total of $2l + 1$ different values (including $l = 0$).

These relations are often represented by a vector diagram. The student is doubtless familiar with the interpretation of angular momentum as a vector quantity. When a rigid body is rotating about a fixed point, like a top about its point, there will be a certain axis through the fixed point about which the angular momentum of the body is a maximum. About any other axis, inclined at an angle θ to the line of maximum angular momentum, the angular maximum is equal to the maximum value multiplied by $\cos \theta$. The angular momentum can be represented by a vector drawn in the direction of the axis for a maximum; the length of this vector can be made to represent the magnitude of the maximum angular momentum, and its direction, indicated by an arrowhead, is commonly taken in the direction in which a right-handed screw would advance along the axis while turning about it in the same direction as the rotating body. The angular momentum about any other axis is then represented by the corresponding component of this vector. The vector may or may not coincide with the instantaneous axis of rotation of the body. A similar treatment can be given to a moving set of disconnected particles.

The various values of angular momentum associated with the quantum states can be pictured, therefore, by imagining a vector of length l , or $lh/2\pi$, placed in any one of a discrete set of positions relative to the polar axis, so that its component along this axis is of magnitude λ or $\lambda h/2\pi$ units (Cf. Fig. 88). No significance is to be attached to the side of the axis on which the l -vector is drawn, however; the components of the angular momentum perpendicular to the axis are not fixed at all. Thus there is for the electron no definite direction for its resultant angular momentum as a vector, contrary to the situation in classical theory.

The square of the resultant angular momentum, on the other hand, is fixed, according to wave mechanics, at the rather curious value of $l(l+1)(h/2\pi)^2$. This implies a resultant angular momentum of magnitude $\sqrt{l(l+1)} h/2\pi$, not $lh/2\pi$ as might have been anticipated. Sometimes the vector is drawn of a length to correspond to the value $\sqrt{l(l+1)} h/2\pi$. The simpler length l is adequate, however, for pictorial purposes; perhaps it should be interpreted as representing the quantum number l rather than the angular momentum itself.

These relationships are sometimes referred to as *space quantization*.

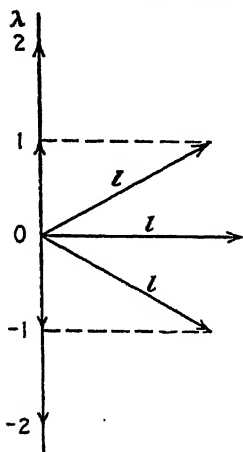


FIG. 88.—Illustrating the quantization of angular momentum for $l = 2$.

The axis of polars can, in general, be drawn in any desired direction; thus the set of quantum states is in some degree a matter of arbitrary choice. (An analogous case in ordinary mechanical vibrations is the fact that, in the case of a circular drumhead vibrating with a single nodal line and emitting one of its overtones, the nodal line may lie along *any* diameter. Degeneracy occurs here, too.) When a magnetic field is present, however, the axis must be taken in the direction of this field.

No such interpretation can be given for the number n , which is often called the *principal* quantum number.

(c) *Probability Density and Charge-cloud Density*.—The probability density corresponding to one of the wave functions $\psi_{n\lambda}$ is easily seen from Eq. (185) to be

$$|\psi_{n\lambda}|^2 = |C_\lambda|^2 [P_\lambda^2(\cos \theta)]^2 [R_{n\lambda}(r)]^2. \quad (187)$$

There is little suggestion here of anything like *orbital motion*. The quantity $|\psi_{n\lambda}|^2 d\tau$ represents the probability that, if the electron were located experimentally at any instant, it would be found in the element of volume $d\tau$ (Sec. 120). It would not be possible to follow it in an orbital motion around the nucleus by means of a succession of such observations, however, as the astronomers follow the planets around the sun; for one observation with a γ -ray microscope of sufficient resolving power would suffice, because of the Compton effect, to knock the electron entirely out of the atom. Often it is convenient to imagine the electronic charge e distributed in space as a sort of *charge cloud*, with a density

$$\eta = e|\psi|^2.$$

Many effects of the atom on its surroundings are approximately the same as if the atom actually contained a distribution of charge of density η , acting according to the ordinary laws of electrostatics. In any case, this furnishes a convenient means of visualizing the probability distribution.

From Eq. (187) it is evident that, when the electron is in one of its quantum states, the probability density or charge-cloud density is symmetrical about the axis of polars. The variation with θ may be large; but this is of minor interest. For this reason it is more illuminating to consider only the *average radial* variation in the density. Let $P_r dr$ denote the total amount of probability, or $\eta_r dr$ the numerical amount of charge, contained between two spheres of radii r and $r + dr$ drawn about the nucleus as a center. Then it is easily seen that, if R_{ni} is separately normalized so that

$$\int_0^\infty r^2 R_{ni}^2 dr = 1$$

(C_n being then so chosen that

$$|C_n|^2 \iint [P_r (\cos \theta)]^2 \sin \theta d\theta d\phi = 1),$$

then

$$P_r = r^2 R_{ni}^2;$$

and then

$$\eta_r = eP_r = er^2 R_{ni}^2.$$

(d) *Further Details Concerning the Wave Functions.*—As examples, the following mathematical details may be given concerning the states of lowest energy, written in terms of

$$a_0 = \frac{h^2}{4\pi^2 me^2} = 5.27 \times 10^{-9} \text{ cm.};$$

$$n = 1, \quad l = 0, \quad \lambda = 0: \quad \psi_{100} = \frac{1}{\sqrt{4\pi}} R_{10}, \quad R_{10} = 2 \left(\frac{Z}{a_0} \right)^{3/2} e^{-Zr/a_0};$$

$$n = 2, \quad l = 0, \quad \lambda = 0: \quad \psi_{200} = \frac{1}{\sqrt{4\pi}} R_{20},$$

$$R_{20} = \frac{1}{\sqrt{2}} \left(\frac{Z}{a_0} \right)^{3/2} \left(1 - \frac{Zr}{2a_0} \right) e^{-\frac{Zr}{2a_0}};$$

$$l = 1, \quad \lambda = 0: \quad \psi_{210} = \sqrt{\frac{3}{4\pi}} \cos \theta \cdot R_{21},$$

$$R_{21} = \frac{1}{2\sqrt{6}} \left(\frac{Z}{a_0} \right)^{3/2} \frac{Zr}{a_0} e^{-\frac{Zr}{2a_0}};$$

$$\lambda = \pm 1: \psi_{21, \pm 1} = \sqrt{\frac{3}{8\pi}} e^{\pm i\varphi} \sin \theta \cdot R_{21};$$

for $n = 3$,

$$R_{30} = \frac{2}{3\sqrt{3}} \left(\frac{Z}{a_0} \right)^{3/2} \left[1 - \frac{2}{3} \frac{Zr}{a_0} + \frac{2}{27} \left(\frac{Zr}{a_0} \right)^2 \right] e^{-\frac{Zr}{3a_0}},$$

$$R_{31} = \frac{4}{27} \sqrt{\frac{2}{3}} \left(\frac{Z}{a_0} \right)^{3/2} \left(1 - \frac{1}{6} \frac{Zr}{a_0} \right) \frac{Zr}{a_0} e^{-\frac{Zr}{3a_0}},$$

$$R_{32} = \frac{4}{81} \sqrt{\frac{1}{30}} \left(\frac{Z}{a_0} \right)^{3/2} e^{-\frac{Zr}{3a_0}}.$$

The constant a_0 is the same number that represented the radius of the smallest orbit for hydrogen [cf. Eq. (135)] in the Bohr theory. For $Z = 1$, these formulas refer to hydrogen; for $Z = 2$, to ionized helium; and so on. In Fig. 89 are plotted not R_{nl} but rR_{nl} ; and in Fig. 90 are shown the corresponding curves for P_r or η_r , which is proportional to $r^2 R_{nl}^2$. The areas under the latter curves are all equal (except for change of scale). The curves are labeled in the notation of

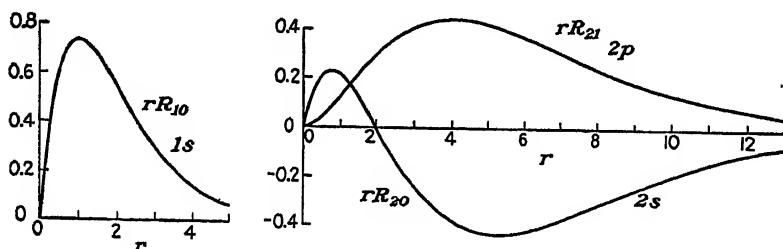


Fig. 89.—The radial function rR_{nl} for several quantum states of the hydrogen atom.

the spectroscopists, the letters s, p, d, f, \dots being used to indicate $l = 0, 1, 2, 3, \dots$, preceded by a number giving the value of n . The meaning of these letters needs to be learned:

$$\begin{array}{ccccccc} l = & 0 & 1 & 2 & 3 & 4 & \dots \\ & s & p & d & f & g & \dots \end{array}$$

Although there is nothing in the wave functions that suggests the former Bohr orbits, a certain correspondence between them can be traced. The orbit had one more unit of angular momentum than has the corresponding state, *viz.*, k units of $h/2\pi$ where $k = l + 1$. In the normal or lowest energy state (or "ground state"), there is no angular momentum, according to wave mechanics, as against 1 unit in the Bohr theory. States with $l = n - 1$ correspond to circular orbits, the other states to elliptical orbits of various eccentricities; the s state, $l = 0$, corresponds to the ellipse of maximum eccentricity for given n . The various values of λ correspond to

various inclinations of the plane of the orbit to the axis, the plane being parallel to the axis for $\lambda = 0$ and perpendicular to it for $\lambda = \pm l$.

A further connection with the Bohr orbits is that, in states with $l = n - 1$, the maximum radial probability density P_r or charge density η_r occurs at a value of r equal to the radius of the corresponding Bohr circle. In these states with $l = n - 1$, as is evident from Fig. 90, the density exhibits the greatest degree of crowding into a single large hump; in the s states ($l = 0$), it is most widely distributed

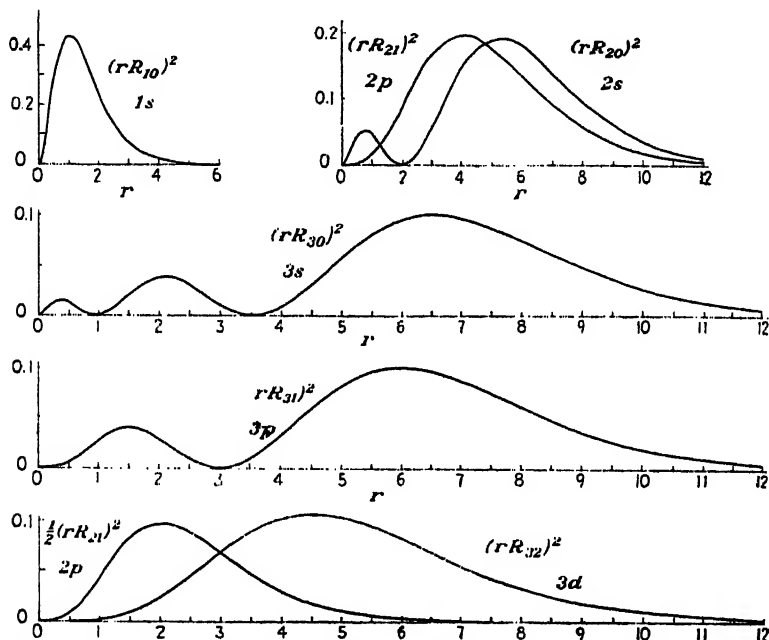


FIG. 90.—The radial probability density $P_r = r^2 R_{nl}^2$ for several quantum states of the hydrogen atom (one curve being drawn twice).

From the formulas, it is evident that for s states ψ is actually finite at the nucleus ($r = 0$). This fact tends to give peculiar properties to s states.

As n increases, the curves for R_{nl} , or for P_r or η_r , spread out more and more, just as did the Bohr orbits; the atom swells, so to speak. If Z increases, on the other hand, it is evident from the formulas that the atom shrinks in the inverse ratio of Z .

As a final remark, it may be stated that the motion of the nucleus introduces exactly the same effect as we found it to do on the Bohr theory: the electronic mass m is merely replaced in all formulas by the reduced mass (Sec. 103)

$$m' = \frac{mM}{m + M}$$

(M = mass of nucleus).

125. Emission and Absorption of Radiation.—One of the most important properties of atoms and molecules is their ability to emit or absorb radiant energy. We cannot take up here the mathematical theory of this process; but we may describe the quantities that are calculated theoretically with the help of wave mechanics, and a word or two may be added in regard to the method of procedure.

The most comprehensive method of treating radiation problems is to resolve the electromagnetic field inside a large box into standing-wave systems or harmonic modes of oscillation, just as we did in Sec. 87 for the purpose of deducing the laws for black-body radiation. Each mode of oscillation of the field is then treated as a harmonic oscillator, the wave-mechanical theory described in Sec. 123 being used for it; and radiation problems are solved by calculating the effect of the interaction between these field oscillators and any atoms or molecules that may be immersed in the field.

(a) *Einstein's A and B Coefficients. Mean Life.*—Problems in the emission of radiation center about the determination of constants known as the *probabilities for spontaneous radiative transition* or *Einstein's A coefficients*. They can be defined as follows. Suppose an atom is in its quantum state number n and that state number j has lower energy. Then, in any short interval of time dt , there will be a certain probability that the atom will jump from state n into state j , the difference in energy between the two states being emitted as a photon of radiation. This probability is denoted by $A_{nj} dt$. Or, we can say that out of a large number N of atoms in state n , $NA_{nj} dt$ will jump into state j during the interval of time dt . Thus A_{nj} represents the probability per second of a spontaneous jump from state n to state j accompanied by the emission of a photon.

The A coefficients are connected with the *mean life* of an atom in a given quantum state, an idea that is often useful. Suppose that N_0 atoms start in state n at time $t = 0$, and that after a time t all have jumped into other states except N of them. Then, during the next element dt of time,

$$N \sum_j' A_{nj} dt$$

of them will jump, where \sum_j' denotes a summation over all states of lower energy only. Thus, during dt , N changes by dN where

$$dN = -N\gamma dt, \quad \gamma = \sum_j' A_{nj},$$

so that, after integrating,

$$N = N_0 e^{-\gamma t};$$

here we have chosen the constant of integration so that $N = N_0$ when $t = 0$. (Cf. Fig. 91.) Let us calculate the average time spent by all N_0 atoms in state n , which is called the *mean life* of an atom in that state; denote it by τ_n . Since $-dN$ or $N\gamma dt$ of them jump after a time t spent in state number n , the average time spent in that state is

$$\tau_n = \frac{1}{N_0} \sum t(-dN) = \frac{1}{N_0} \int_0^\infty N\gamma t dt,$$

or, after inserting the value of N and integrating by parts,

$$\tau_n = -te^{-\gamma t} \Big|_{t=0}^\infty + \int_0^\infty e^{-\gamma t} dt.$$

But $te^{-\gamma t} \rightarrow 0$ as $t \rightarrow 0$ or as $t \rightarrow \infty$; and $\int_0^\infty e^{-\gamma t} dt = 1/\gamma$. Hence

$$\tau_n = \frac{1}{\gamma} = \left(\sum_j' A_{nj}\right)^{-1}. \quad (188)$$

For atomic states involved in the emission of visible or ultraviolet radiation, τ_n is very roughly of the order of 10^{-8} sec.; in the X-ray region, it is much smaller.

For the *absorption* of radiation, a corresponding coefficient B is defined as follows. Consider, as before, two quantum states for an atom, numbers n and j , and let ν_{nj} be the frequency of the light emitted by the atom when it jumps spontaneously from n to j . Suppose the atom is bathed by radiation with a spectral energy density

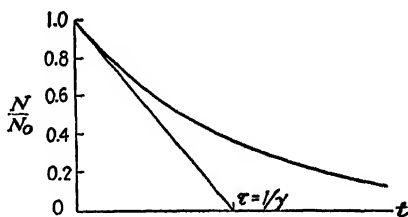


FIG. 91.—Illustrating exponential decrease by radiative transitions. $\tau = \tau_n =$ mean life.

equal to u_ν in the neighborhood of the frequency ν_{nj} ; i.e., there are $u_\nu d\nu$ ergs of radiant energy per cm.³ in the frequency range $d\nu$. Then, if the atom is in state j , there is, during each element of time dt , a certain probability that the atom will absorb a photon of frequency ν_{nj} and jump up to state n ; this probability will be proportional to u_ν , so we may write for it $B_{nj}u_\nu dt$, in terms of a new constant B_{nj} . If, on the other hand, the atom is already in state n , the theory indicates that its probability of jumping into state j is increased by the

presence of the radiation from the value $A_{nj} dt$ to $(A_{nj} + B_{nj}u_r) dt$. The term $B_{nj}u_r dt$ is said to represent *induced emission*. We can say that the radiation induces the same probability of jumping either way between the two states, the probability per unit time being $B_{nj}u_r$. (The same conclusion follows from classical theory; radiation falling on a vibrating oscillator causes it either to gain or to lose energy depending upon the phase relation between the waves and the vibration of the oscillator.) The quantity B_{nj} , often called "Einstein's B coefficient," might be called the *probability coefficient for an induced radiative transition*.

Although A_{nj} and B_{nj} seem to be quite distinct from each other, theory indicates that a simple general relationship should exist between them, *viz.*:

$$B_{nj} = \frac{c^3}{8\pi\hbar\nu_{nj}^3} A_{nj}, \quad (189)$$

where h is Planck's constant and c the speed of light. Thus, the ratio B_{nj}/A_{nj} is the same for all radiating systems. It follows that for different atoms or molecules the ability to absorb and the ability to emit radiation of any given frequency are proportional to each other. This constitutes an extension to gases of Kirchhoff's law (Sec. 76).

(b) *Einstein's Deduction of Planck's Law*.—The coefficients A and B were originally introduced by Einstein in giving a deduction of Planck's law for thermal radiation based upon the processes of emission and absorption.¹ The proof is very simple and is interesting.

In an isothermal enclosure, let there be N_n molecules in state n and N_j molecules in state j , and let W_n, W_j denote the molecular energies for these two states. Then, when thermal equilibrium obtains, according to the Boltzmann formula [Eq. (148a) in Sec. 110],

$$N_n = C e^{-W_n/kT}, \quad N_j = C e^{-W_j/kT},$$

C being a constant of proportionality. The existence of equilibrium requires that as many atoms jump in a second from n to j , with the emission of a photon, as jump from j to n , accompanied by the absorption of a photon. Hence, in terms of the probability coefficients just defined,

$$N_n (A_{nj} + B_{nj}u_r) = N_j B_{jn} u_r;$$

whence

$$u_r = \frac{A_{nj}}{B_{nj}} \frac{1}{(N_j/N_n) - 1}.$$

But, by the preceding equations,

¹ EINSTEIN, *Phys. Zeits.*, vol. 18, p. 121 (1917).

$$\frac{N_j}{N_n} = e^{h\nu/kT}, \quad \nu = \frac{W_n - W_j}{h}.$$

Hence, using Eq. (189) with ν_{nj} replaced by ν , we obtain

$$u_\nu = \frac{8\pi h \nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1}.$$

This is readily shown to represent Planck's radiation law expressed in terms of frequency instead of wave length. For the frequency ν and wave length λ are related to the speed of light c by the equation $\nu\lambda = c$; hence, as ν and λ vary, $\nu d\lambda + \lambda d\nu = 0$. If we omit the negative sign, taking $d\lambda$ and $d\nu$ both to be positive, we may write

$$\nu d\lambda = \lambda d\nu.$$

Now the same energy which is represented here by $u_\nu d\nu$ was represented by $\psi_\lambda d\lambda$ in Chap. VI. Hence,

$$\psi_\lambda d\lambda = u_\nu d\nu.$$

From these two equations we have

$$\psi_\lambda = \frac{\nu u_\nu}{\lambda}.$$

Inserting here the value just obtained for u_ν and then $\nu = c/\lambda$, we have

$$\psi_\lambda = \frac{8\pi ch}{\lambda^5} \frac{1}{e^{ch/\lambda kT} - 1},$$

which agrees with Eq. (121) in Sec. 93.

(c) *Calculation of Transition Probabilities.*—When theoretical values of the transition probabilities are calculated by the method indicated above, they are found to depend upon certain quantities called *matrix components* between the wave functions of the two states that are involved in the transition.

The most important term in the expression that is obtained for A_{nj} from wave mechanics, for an atom containing 1 electron, is as follows:

$$A_{nj} = \frac{64\pi^4 e^2 \nu_{nj}^3}{3hc^3} (|x_{nj}|^2 + |y_{nj}|^2 + |z_{nj}|^2). \quad (190)$$

Here e = electronic charge, h = Planck's constant, c = speed of light, and ν_{nj} is the frequency of the photon that is emitted as the electron jumps from state n to state j ; x, y, z are the coordinates of the electron and

$$x_{nj} = \iiint x \psi_n^* \psi_j dx dy dz,$$

where ψ_n, ψ_j are the wave functions for the two states, y_{nj} and z_{nj} being given by similar integrals with x replaced by y or z , respectively. The numbers x_{nj}, y_{nj}, z_{nj} are called *matrix components* of x, y , or z , respectively, between the functions ψ_n and ψ_j . They vary according to the values of the two indices, n and j , and the possible values of each of them might be written as a matrix with n numbering the rows and j the columns. ψ_n^* denotes the complex conjugate of ψ_n .

The expression just written for A_{nj} may be compared with that for the emission of radiation by a classical oscillator. In Eq. (40b) in Sec. 38, for the energy radiated per second by a classical harmonic oscillator, consisting of a charge q vibrating in linear harmonic motion with amplitude a , we obtained

$$\frac{16\pi^4\nu^4q^2a^2}{3c^3} \text{ ergs.}$$

If we divide this expression by the amount of energy in a photon, $h\nu$, we have

$$\frac{16\pi^4\nu^3q^2a^2}{3hc^3}$$

as the number of photons of energy emitted by the oscillator per second. This agrees with the value given in Eq. (190) for A_{nj} , which is the probability per second of the emission of a photon, provided we drop y_{nj} and z_{nj} , so as to have an expression referring to one-dimensional motion, and assume that $\nu = \nu_{nj}$ and

$$a = 2x_{nj}.$$

Because of such correspondences with classical theory the term in A_{nj} that is given in (190) is said to refer to *dipole radiation* by the atom. The average rate of emission of dipole radiation by atoms in jumping from state n to state j is the same as if the atoms contained classical oscillators of the corresponding frequency vibrating with amplitudes equal to twice the matrix components of the electronic coordinates between the two states.

There are also other less important terms in the complete expression for A_{nj} . The dipole term, given in Eq. (190), is usually the principal term, unless it actually vanishes. If the dipole terms vanish, the transition from level n to level j cannot occur by dipole emission of radiation. Such transitions are said to be "forbidden" so far as dipole emission is concerned. "Selection rules" stating which transitions are forbidden are of great importance in spectroscopy. A small probability of transition may exist, however, owing to the other terms

in A_{nj} . The next most important one is called the "quadrupole" term.

126. Relativistic Effects and Electron Spin.—The theory described so far has been a nonrelativistic one suited to cases of small energy only. Like Newtonian mechanics, it requires modification to bring it into harmony with the principles of relativity. We cannot say much here about the relativistic form of wave mechanics; but it happens that its principal new feature can readily be described in terms of a physical picture. This feature was, in fact, discovered before the advent of wave mechanics.

In 1925, Uhlenbeck and Goudsmit pointed out¹ that certain features in atomic spectra could be explained if it was assumed that the electron "spins," or rotates about an axis through its center of mass, and that it has both *angular momentum* and a *magnetic moment* associated with this rotation. The angular momentum is constant in magnitude, but because of it the electron possesses two *internal* quantum states. These states can be so chosen, with reference to any axis, that, when the electron is in one of these quantum states, it has internal angular momentum about the axis equal to $\frac{1}{2}$ unit or $\hbar/4\pi$ in one direction; whereas, when it is in the other state it has an equal amount of angular momentum in the opposite direction. In an atom the angular momentum of spin is to be added to that due to orbital motion. The interaction of the associated magnetic moment with electric or magnetic fields, or with the magnetic moments of other electrons, then modifies the atomic energy levels. Examples of such effects will be described in the next chapter.

An astonishing new turn was given to the theory by Dirac, in 1928.² Dirac showed that the most natural way to bring wave-mechanical theory into harmony with the theory of relativity is to adopt quite a different wave equation and that, when this is done, the new equation leads automatically to effects equivalent to those deduced from electron spin. It need not be *assumed* that the electron is spinning or turning on its axis. This theory of Dirac's, although not free from objection, is the most satisfactory one that we have today. According to it, the electron does behave *as if* it had an internal angular momentum of the sort just described and an associated magnetic moment. Furthermore, if a wave packet is formed representing an electron, there is in it, in general, a closed current or eddy of

¹ UHLENBECK and GOUDSMIT, *Naturwiss.*, vol. 13, p. 593 (1925); *Nature*, vol. 117, p. 264 (1926).

² DIRAC, *Roy. Soc., Proc.*, vol. 117, p. 610; vol. 118, p. 351 (1928). Cf. also DARWIN, *Roy. Soc., Proc.*, vol. 118, p. 654 (1928).

probability that can be regarded as analogous to an actual motion of spin.

The complete Dirac theory is seldom used in spectroscopic work, however. For most purposes, it is sufficient to use the conception of electron spin and to introduce certain additional terms into Schrödinger's wave equation. Even in the absence of these added terms, the introduction of the spin doubles the number of the quantum states. Thus for the hydrogen atom we have wave functions that can be denoted by

$$\psi_{n\lambda\mu}$$

in terms of four, instead of three, quantum numbers, $n\lambda$ and μ . (There is really a fifth quantum number also, $s = \frac{1}{2}$, which is associated with the magnitude of the spin momentum; but, since s never changes, it need not be indicated in writing the wave functions.) The function $\psi_{n\lambda\mu}$ is the same function of the space coordinates as in Eq. (185) above, but we now associate with it a definite spin, as indicated by the value assigned to μ ; if $\mu = \frac{1}{2}$, there is a spin angular momentum equal to $\hbar/4\pi$ about the axis of polar coordinates, whereas, if $\mu = -\frac{1}{2}$, this momentum has the value $-\hbar/4\pi$. The total momentum about the axis, due to both spin and orbital motion, is then $(\lambda + \mu)\hbar/2\pi$.

When, however, the additional spin terms are added in the wave equation, although there are just as many wave functions as there are in the absence of these terms, the functions are somewhat different. The degeneracy is also partly removed, the various functions belonging to a given value of n being associated now, in part, with slightly different energies. We shall return to this topic in a later section (Sec. 148).

CHAPTER VIII

ATOMIC STRUCTURE AND OPTICAL SPECTRA

In the discussion of quantum theory in Chap. VI, we considered only atoms containing a single electron. We shall now attempt to explain along similar lines the properties of the numerous kinds of atoms (and molecules) that contain 2 or more electrons.

The Rutherford-Bohr model of the atom was built out of raw material which grew in the field of physics. Originally, however, the atom was a child of chemistry. The ultimate atom model must explain and agree with the facts of chemistry, which are just as cogent as are the facts of physics. In part of the present chapter, we shall deal with the explanation of the periodic table of the elements. In another part, we shall return to the study of spectra. The two topics are closely related; for wave mechanics furnishes the key to the theoretical understanding of both subjects. The mathematical theory encounters such great difficulties, however, that empirical evidence from the facts of spectroscopy has to be invoked in order to interpret some of the finer details of atomic structure which are important to the chemist.

Before taking up either of these topics we shall first describe briefly the manner in which wave mechanics has been extended in order to cover systems of 2 or more electrons.

WAVE MECHANICS OF MANY-ELECTRON SYSTEMS

127. Wave-mechanical Fundamentals.—In Secs. 119 to 124 and 126, we have described the wave-mechanical theory of a single electron in a field of force. Familiarity with this theory will be assumed. When more than 1 electron is present in the atom, there is still a single *probability amplitude* Ψ , and a single *wave equation* for it; but both Ψ and the wave equation now contain the coordinates and symbols denoting the spins of *all* the electrons. Since our purpose is not to develop the mathematical theory in detail but only to describe some of its features and to cite the principal results to which it leads, we shall not write down the complete wave equation but shall only describe some of its terms. Only the *stationary states* of the atom will be discussed, for any one of which we can write

$$\Psi = e^{-\frac{2\pi i}{h} W t} \psi,$$

in which W is the energy of the atom, t , the time, h , Planck's constant, and ψ , a function which does not contain t (cf. Sec. 122).

The wave equation contains a term of the form $V\psi$, but the potential-energy function V now includes not only the potential energy of all the electrons due to the field of the nucleus but also their mutual potential energy due to the repulsion of each one by all of the others. With this preliminary remark, the principal groups of terms in the wave equation (or amplitude equation) for ψ may be listed as follows:

(1) terms containing second derivatives, like those in the equation for 1 electron;

(2) terms due to the nuclear field, which can be written as the sum

$$-e^2 Z \left(\sum_r \frac{1}{r_r} \right) \psi,$$

where e = electronic charge, Z = atomic number, $1/r_r$ = distance of the r th electron from the nucleus, and the sum extends over all electrons in the atom;

(3) terms representing the electrostatic interaction of the electrons, which can be written

$$e^2 \left(\sum_{\rho < \tau} \sum \frac{1}{r_{\rho\tau}} \right) \psi,$$

where $r_{\rho\tau}$ is the distance between electrons number ρ and τ expressed in terms of their coordinates, so that $e^2/r_{\rho\tau}$ is the mutual potential energy of this pair, and the sum extends over all pairs that can be formed out of the whole group of electrons;

(4) spin-orbit terms representing an interaction between the spin magnetic moment of each electron and its own orbital motion.

In addition, there are a few terms of minor importance representing mixed spin effects and certain other relativistic effects, but with these we need not concern ourselves.

In the terms of type (3), the coordinates of different electrons occur combined together in such fashion as to make the solution of the wave equation very difficult. For this reason, little progress has been made except by means of a method of approximation known as perturbation theory. In using this method, the interaction of the electrons with each other is at first omitted entirely; *i.e.*, each electron is assumed to move in the field of the nucleus alone, just as if the other electrons were not there. Or, still better, the field of the nucleus, which is often

called a *Coulomb field*, is replaced by a *modified central field*¹ which includes an allowance for the *average* effect of the electrons on one another.

The allowance for the interaction of the electrons may be visualized as follows. In the case of 1 electron, as explained in Sec. 120, $|\psi|^2 d\tau$ was interpreted as the probability of finding the electron, and also its total electrical charge, in the element of volume $d\tau$, ψ being given the value that it has at some point in $d\tau$. This rule could be visualized by imagining the electronic charge e distributed continuously in space with a density $e|\psi|^2$; the *fraction* of the charge thus assigned to $d\tau$ would then be the same as the probability of finding the electron in $d\tau$. In a similar way, the charges on all the electrons in a many-electron atom may be imagined spread out in a continuous "*charge cloud*" whose density is determined in a certain manner by ψ . The conception is an artificial one, but it is useful.

Such a distributed charge, according to the ordinary laws of electrostatics, would produce a negative electrostatic potential. The modified central field that is used in perturbation theory can now be regarded as formed by superposing the field of potential due to the electronic charge cloud upon the Coulomb field² due to the nucleus. The result is a field the potential of which becomes indistinguishable from the nuclear potential Ze/r (e = numerical electronic charge, Z = atomic number) near the nucleus but falls off more rapidly at greater distances. The electronic charge cloud is often thought of as "screening off" part of the nuclear field. Outside of the atom, the potential in the central field drops rapidly to zero if the atom is electrically neutral, otherwise it passes into the Coulomb field of a charge equal to the net charge on the atom. Strictly speaking, however, the central field formed exactly as described is suitable only for treating the deflection of an external electron by the atom, as in calculating the scattering of cathode rays. In dealing with the motion of one of the electrons in the atom, only the potential due to the charge cloud of the others, or the fraction $(N - 1)/N$ of the potential due to all N electrons in the atom, should be superposed upon the nuclear field. Several good methods have been invented for obtaining appropriate central fields without a prior knowledge of ψ .*

¹ By a "central field" is meant one in which the potential energy of an electron depends only on its distance from a fixed point, so that the force on it, in classical theory, is also determined by its distance from the point and is directed toward or away from the latter.

² A "Coulomb field" is like that which is due to a point charge.

* Cf. CONDON, E. U., and G. H. SHORTLEY, "Theory of Atomic Spectra," 1935.

In Fig. 92 is shown such a central field for the atom of argon, and also the charge-cloud density, as obtained using the Hartree consistent-field method with allowance for the effects of exchange.¹ The curve labeled Z_p shows, in a certain manner, the potential energy U that an additional electron would have at various distances r from the nucleus. The potential energy itself is not plotted, however, because it varies so rapidly with r , but, rather, the number of units of charge Z_p on an equivalent nucleus; i.e., the ordinate of the curve at any given value of r shows the atomic number of a nucleus which, at that particular distance, would give to an electron a potential energy equal to U , so that $U = -eZ_p/r$. The other curve is a plot of η_r , $\eta_r dr$ being

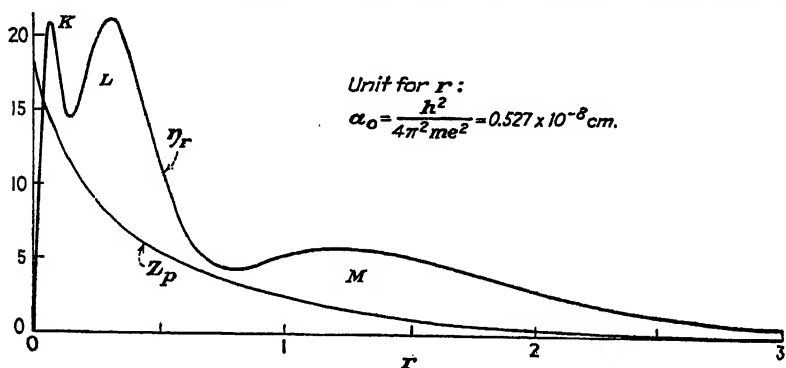


FIG. 92.—The Hartree field and charge-cloud density for argon. The potential energy of an additional electron inside the argon atom would be $-eZ_p/r$ (e = numerical electronic charge, r = distance from the nucleus); the amount of electronic charge between r and $r + dr$ in terms of e as a unit is $\eta_r dr$. K , L , M indicate the approximate locations of the K , L , and M shells.

the amount of electronic charge between distances r and $r + dr$ from the nucleus. Thus, Z_p varies from $Z_p = Z = 18$ at $r = 0$ to $Z_p = 0$ at the periphery of the atom. The maxima in η_r correspond to the various electronic shells, as is indicated in the figure.

128. The Zero-order Approximation: Spin Energy Omitted.—

Having selected a suitable central field, the next step is to ascertain the wave functions corresponding to the various possible *quantum states for a single electron* in this field. For this purpose, it is often convenient at first to omit the spin terms from the wave equation. When this is done, the 1-electron wave functions can be so chosen that each one contains the same sort of angle function as occurs in the wave functions for the 1-electron atom as written above (Sec. 124). The radial factor in the wave function, or function of the distance r from the nucleus, is different here, however, and can be found only by the

¹ HARTREE and HARTREE, *Roy. Soc., Proc.*, vol. 166, p. 450 (1938).

process of numerical integration. Furthermore, although the spin terms have been omitted from the wave equation, spin must still be taken into account in drawing up the list of 1-electron quantum states. The angular momentum due to spin, about any chosen axis, can have either of two possible values, $\hbar/4\pi$ or $-\hbar/4\pi$. This is indicated by adding a fourth quantum number, μ , which may be equal to either $\frac{1}{2}$ or $-\frac{1}{2}$, denoting the angular momentum about the axis due to spin, expressed in units of $\hbar/2\pi$. (Often m_s is used instead of μ .)

Each quantum state or wave function for 1 electron is then characterized by four quantum numbers, n, l, λ, μ . As in Sec. 124, n, l , and λ are any integers such that

$$|\lambda| \leq l < n,$$

l and n being, therefore, always positive. (Often m or m_l is used instead of λ .) There is also a fifth quantum number s , always equal to $\frac{1}{2}$, but, since this number never changes, it may be ignored for the time being. The radial factor $R_{nl}(r)$ and the allowed values of the energy are now found to vary with l as well as with n . Thus, part of the degeneracy found with a field of Coulomb type was of the nature of a mathematical accident and does not occur with more general types of field.

An electron put into the field can exist there in a state represented by any one of these 1-electron wave functions. If more electrons are added, and if we continue to ignore the part of their electrostatic interaction that is not effectively included in the central field, it would be expected that each one could be represented by any one of the 1-electron functions. According to the Bohr picture, each electron would then be moving in the field along some orbit independently of the others. In wave mechanics, however, the electrons are not quite independent even when their electrostatic repulsion is ignored. *No two electrons can be in the same 1-electron quantum state represented by the same wave function with the same values of n, l, λ, μ .* This restriction was originally laid down by Pauli. It can be derived from a certain sort of symmetry that the wave function for the entire atom must possess,¹ but no reason can be given (as yet) for requiring this type of symmetry in ψ , except the fact that, if it is required, the results of the theory are thereby brought into agreement with experi-

¹ The atomic wave function ψ must be antisymmetric in the electrons, *i.e.*, ψ must be such a function that it reverses its sign, without other change, when the coordinates and spins of any 2 electrons are interchanged in it. For example, $\sin(x_1 - x_2)$ is antisymmetric in x_1 and x_2 , but $\sin x_1 \cos x_2$ or $\sin(x_1 + x_2)$ are not.

ment. The restriction thus imposed upon the admissible solutions of the wave equation is known as the *exclusion principle*. It plays an important role in atomic theory.

The first step in obtaining a wave function corresponding to a certain *quantum state for the atom as a whole* is to assign a 1-electron wave function to each electron. A first or "zero-order" approximation to a wave function ψ for the atom as a whole is then constructed out of these 1-electron functions in a manner that need not be described here. This atomic function, or the corresponding atomic quantum state, can be regarded as characterized by the sets of quantum numbers specifying the chosen 1-electron wave functions, *e.g.*, for N electrons, by the following N set of numbers, all sets differing in at least one number:

$$n_1 l_1 \lambda_1 \mu_1, \quad n_2 l_2 \lambda_2 \mu_2, \quad \dots, \quad n_N l_N \lambda_N \mu_N. \quad (A)$$

Even in this first stage of perturbation theory, however, it makes no sense to ask *which* electron is represented by which of the 1-electron wave functions. Here is another point of contrast with classical ideas. The atomic wave function is constructed out of the 1-electron functions in such a way that the coordinates of each electron occur once in each of the N 1-electron functions (just as in the function, $\sin ax \sin by - \sin ay \sin bx$, x and y both occur once in each of the two sine functions characterized by a and b). Thus, according to wave mechanics, the electrons do not seem to possess the same degree of individuality that we attach, say, to separate grains of sand. Sometimes the electrons are imagined to take turns moving in orbits corresponding to the various 1-electron functions; such an assumption is harmless, but there is nothing in the theory to suggest such a periodic exchange of positions.

For the same reason, the total charge-cloud density at a given point is not just the sum of the charge densities corresponding to the separate electronic functions, as classical analogies would lead us to expect. Its integrated value, however, must come out equal to the total charge on all the electrons; and, for some purposes, the finer details concerning the charge cloud, which are connected with the exchange effect mentioned below, can be ignored. As a first approximation, therefore, we may imagine the charge on each electron to be distributed in space in the characteristic manner indicated by its 1-electron wave function.

Associated with each electronic wave function, there is a certain value of the energy of the electron in the central field. Contrary to the situation in the Coulomb field, as was stated above, this energy

depends upon l as well as upon n ; hence, we shall denote it by W_{nl} . The value of W_{nl} represents a first approximate value of the work which would be required to remove an electron whose wave function is characterized by n and l to infinity, thereby ionizing the atom. The energy of the atom as a whole, however, is different from the sum of the energies W_{nl} for all the electrons, essentially because some of the energy is mutual energy and is counted twice when we simply add up the W_{nl} 's.

Since the electronic energies depend only upon n and l (as long as the spin energy is neglected), the quantum numbers n and l furnish a convenient means of classifying the electrons in an atom. For this purpose, however, a notation developed first by the spectroscopists is often more convenient. In this notation small letters are used to denote the l values, according to the following scheme:

$$\begin{array}{cccccccc} & s & p & d & f & g & h & \cdots \\ l = & 0 & 1 & 2 & 3 & 4 & 5 & \end{array}$$

(The reason for the choice of letters is stated in Sec. 142 below.) The number of electrons that have the same n and l is often indicated by a superscript. For example,

$$2s^2 3d$$

denotes 2 electrons having wave functions of the $2s$ type with $n = 2$, $l = 0$ and, of course, with some difference in the unspecified λ and μ values, in order to satisfy the exclusion principle; a third electron has a $3d$ wave function with $n = 3$, $l = 2$ (and unspecified values of λ and μ).

Electrons having the same n are said to belong to the same *shell*. Each shell is then regarded as divided into *subshells* according to the values of l . Any subshell may contain up to

$$2(2l + 1)$$

electrons, but no more. For λ can range from $\lambda = l$ down through zero to $\lambda = -l$, making $2l + 1$ different values for λ ; and, in each case, μ may be either $\frac{1}{2}$ or $-\frac{1}{2}$, whence the factor 2. The numbers of electrons in the various types of closed subshells are, therefore, as follows:

$$\begin{array}{cccccc} s & p & d & f & g & \\ 2 & 6 & 10 & 14 & 18 & \end{array}$$

The maximum number of electrons in a whole *shell*, on the other hand, is

$$2n^2.$$

For, since $0 \leq l < n$, it is

$$\sum_{l=0}^{n-1} 2(2l+1).$$

Here in the series $\sum_{l=0}^{n-1} l = 0 + 1 + 2 + \cdots (n-1)$, there are n terms, and their average value is $[0 + (n-1)]/2 = (n-1)/2$; hence, the sum of this series is $n(n-1)/2$. Furthermore, $\sum_{l=0}^{n-1} 1 = n$.

Thus, $\sum_{l=0}^{n-1} (2l+1) = n(n-1) + n = n^2$.

The shells are often referred to by the capital letters which became attached to them in the early days of X-ray study (Sec. 190). These designations are shown below, also, in part, in Table I, in which the numbers and notation for the first three shells are summarized:

$n =$	1	2	3	4	5	6	7
X-ray designation:	<i>K</i>	<i>L</i>	<i>M</i>	<i>N</i>	<i>O</i>	<i>P</i>	<i>Q</i>

TABLE I.—ELECTRON SHELLS AND SUBSHELLS

Shell.....	<i>K</i>	<i>L</i>		<i>M</i>		
n	1	2		3		
Subshell, l	0	0	1	0	1	2
Letter designation.....	<i>s</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>d</i>
Number of electrons in subshell or shell. }	2	2	6	2	6	10
	2	8		18		

A more complete list is contained in Table III in Sec. 137. As an illustration of the various combinations of quantum numbers for the 1-electron wave functions, those for the first two shells are listed in Table II.

TABLE II.—ONE-ELECTRON QUANTUM NUMBERS IN THE *K* AND *L* SHELLS

Shell	n	l	λ (or m_l)	μ (or m_s)
<i>K</i>	1	0	0	$+\frac{1}{2}$
	1	0	0	$-\frac{1}{2}$
<i>L</i>	2	0	0	$+\frac{1}{2}$
	2	0	0	$-\frac{1}{2}$
	2	1	1	$+\frac{1}{2}$
	2	1	1	$-\frac{1}{2}$
	2	1	0	$+\frac{1}{2}$
	2	1	0	$-\frac{1}{2}$
	2	1	-1	$+\frac{1}{2}$
	2	1	-1	$-\frac{1}{2}$

129. The More Accurate Theory.—The method of attack just described constitutes only a first approximation. To make it more accurate, two additional sets of terms in the wave equation must be considered, representing the following: (1) the *residual part* of the electronic interaction, due to electrostatic repulsion, over and above the average effect which has been allowed for in choosing the modified central field; and (2) the spin-orbit terms, representing the interaction of the spin magnetic moment of each electron with its own orbital motion.

The effect of restoring these terms in the wave equation is to alter somewhat both the atomic wave functions and the energies of the atomic quantum states. If the spin terms alone are included, it is found that they can be allowed for by modifying the 1-electron functions and the associated 1-electron energies, with the substitution of quantum numbers $n\ l\ j\ m$ for $n\ l\ \lambda\ \mu$; this procedure is sometimes useful and will be described in more detail later (Sec. 153). The electrostatic interaction, however, can be allowed for only by the methods known as *perturbation theory*; and actually the spin-orbit effect is usually treated in this same way. The procedure is analogous to one used by the astronomers; hence, the name. In treating the motion of the planets, the astronomer first assumes each planet to move in a fixed elliptical orbit about the sun, then he calculates the *perturbations* of these orbits caused by the attraction of all the other planets. In atomic theory, the mathematical work is so laborious that it is hardly ever carried beyond the first step.

When the *electrostatic interaction between different electrons* is introduced, it is no longer possible, strictly speaking, to suppose that the atomic wave function is made up out of a single set of 1-electron functions. Several sets of these functions must be combined in certain proportions. One set may predominate in the true atomic wave function, however, and in such cases it is convenient to label the corresponding state of the atom with the electronic quantum numbers that specify the predominant set of 1-electron functions.

To the first order, the effect upon the atomic energy of the electrostatic interaction between the electrons can be regarded as composed of two independent partial effects. One part of the change in the energy can be regarded as the mutual electrostatic energy of the charge clouds of the 2 electrons, due to their mutual repulsion. The effect of this is to raise the atomic energy. The other effect of the interaction arises from that symmetry requirement on the atomic wave function that is responsible for the exclusion principle.

In the preceding discussion, we have not specified definitely whether the *number of electrons present* in the field of the nucleus is

assumed to be equal to the atomic number Z of the nucleus or whether it may be less than Z . The same mode of theoretical attack is, in fact, applicable to both cases. If the number of electrons is equal to Z , we have the ordinary or neutral atom; if it is one less than Z , the atom is singly ionized; if two less, it is doubly ionized; and so on.

Furthermore, in the discussion we have completely neglected the motion of the nucleus. In the case of an isolated atom, the effect of nuclear motion is very small, especially in the heavier atoms. Wave mechanics is also applicable, however, to molecules, and the quantum states of a molecule containing two or more nuclei are greatly affected by the relative motion of the nuclei.

This is about as far as it is feasible to go in discussing the wave mechanics of systems containing many electrons without encountering complicated mathematics. We shall find, however, that a great deal can be understood about the structure of atoms, about the combination of atoms into molecules, and about the spectra emitted by atoms and molecules, on the basis of the ideas that have been presented.

THE PERIODIC TABLE OF THE ELEMENTS

130. Atomic Numbers.—Mention was made, in discussing the Rutherford atom model, of the concept of atomic number. This concept was firmly established by the work of Moseley on X-rays, to be described later. As a result of Moseley's work, the total number of elements up to and including uranium was fixed at 92, and the atomic number of each element was unambiguously assigned. The *periodic table* of the elements, drawn up by the chemists on the basis of chemical facts primarily, was thus established with the assurance that *no unknown* gaps were left in it. The atomic numbers of the various chemical elements are shown in Appendixes I and II, the latter giving the periodic table in its standard form as used by the chemists.

The atomic number of an atom appears, in Rutherford's theory of the scattering of α particles and in Moseley's work, as the numerical ratio of the charge on the nucleus to the charge on the electron. This ratio is numerically equal to the number of electrons surrounding the nucleus when the atom is in its normal, neutral condition. When, as a result of the action of some physical agency, an atom becomes *ionized* with the loss of 1 or more electrons, the number of electrons in it is less than its atomic number. The *chemical properties* of an element, however, are understood to refer to the *neutral atom*. Hence the *chemical properties* of an element can be regarded as depending *either* upon its *atomic number* or upon the *number of electrons* in it when it is electrically neutral.

The problem of the periodic table is to be solved, therefore, by determining somehow the *arrangement of the electrons* in the atoms of the various chemical elements and the physical and chemical properties that follow from this arrangement.

Diagram illustrating Bohr's periodic table of the elements, showing the arrangement of elements by periods and groups, with subshells (s, p, d, f) indicated.

Period	Group I	Group II	Group III	Group IV	Group V	Group VI	Group VII
1	1H						
2	2He	3Li	4Be	5B	6C	7N	8O
3		9F	10Ne	11Na	12Mg	13Al	14Si
4		15P	16S	17Cl	18A	19K	20Ca
5		21Sc	22Ti	23V	24Cr	25Mn	26Fe
6		27Co	28Ni	29Cu	30Zn	31Ga	32Ge
7		33As	34Se	35Br	36Kr	37Rb	38Sr
8		39Y	40Zr	41Nb	42Mo	43Tc	44Ru
9		45Rh	46Pd	47Ag	48Cd	49In	50Sn
10		51Sb	52Te	53I	54Xe	55Cs	56Ba
11		57La	58Ce	59Pr	60Nd	61Pm	62Sm
12		63Eu	64Gd	65Tb	66Dy	67Ho	68Er
13		69Tm	70Yb	71Lu	72Hf	73Ta	74W
14		75Re	76Os	77Ir	78Pt	79Au	80Hg
15		81Tl	82Pb	83Bi	84Po	85At	86Rn
16		87Fr	88Ra	89Ac	90Th	91Pa	92U

Subshells indicated: (1s), (2s), (3s), (4s), (5s), (6s), (7s), (2p), (3p), (4p), (5p), (6p), (7p), (3d), (4d), (5d), (6d), (7d), (4f), (5f), (6f), (7f).

FIG. 93.—Bohr's periodic table of the elements. Complete subshells contain two *s*, six *p*, ten *d*, or fourteen *f* electrons. Irregularities in the progression occur at Cr, Cb, Mo, Ma, Ru, Rh (see Appendix III).

131. Some Features of the Periodic Table.—A bird's-eye view of the chief features to be explained is best gained from an arrangement of the periodic table given by Bohr, which is reproduced in Fig. 93. The most striking feature is undoubtedly the repeated recurrence of a noble or *inert* gas, forming a series of turning points in the progression of the elements. These gases are the elements helium ($Z = 2$), neon (10), argon (18), krypton (36), xenon (54), and radon or

radium emanation (86). It was pointed out by Rydberg, soon after Moseley's work had fixed the atomic numbers of the elements, that the values of the atomic number Z for the inert gases can be expressed by simple numerical series:

$$Z = 2(1^2 + 2^2 + 2^2 + 3^2 + 3^2 + 4^2 + \dots);$$

i.e., $2 \times 1^2 = 2$ (helium), $2(1^2 + 2^2) = 10$ (neon); and so on up to 86. Such a fact must obviously find its explanation somehow in the arrangement of the electrons in the atom. The numerical series suggests that the electrons in these inert gases may be arranged in "layers" of some sort; the third inert gas, argon, for instance, would consist of three layers, the first containing 2×1^2 electrons, the second, 2×2^2 , and the third, 2×2^2 . The factor 2 suggests symmetry of some kind, or some general reason for the occurrence of electrons in closely associated pairs.

The "inert" gases, as the name indicates, show comparatively little tendency for their atoms to associate themselves with other atoms in chemical combination. The elements standing in the table on either side of an inert gas, on the other hand, are strongly active chemically, and have contrasting properties. Those closely *following* an inert gas, like lithium, beryllium, potassium, calcium, and so on, are metallic and strongly electropositive; they readily form *positive* electrolytic ions. Furthermore, their maximum valence in chemical compounds is equal to the number of steps by which they lie beyond the inert gas (e.g., 1 for sodium, 2 for magnesium, 3 for aluminum; sodium forms univalent ions in solution, magnesium, bivalent, aluminum, trivalent). The elements closely *preceding* an inert gas, on the contrary, are electrically nonconducting, perhaps even gaseous, and strongly electronegative; they tend to form *negative* electrolytic ions, alone or in combination with other atoms, and they tend to exhibit chemical valence equal to the number of steps by which they precede the inert gas in the table. An element of either of these two kinds scarcely combines at all with another element of the same kind; but an element closely preceding an inert gas in the table combines readily with any other element closely following an inert gas.

These facts invite the conclusion that there is something very peculiar about the arrangement of the electrons in an atom of an inert gas. If we make the reasonable assumption that the chemical activity of an atom is conditioned somehow upon the magnitude of its external electric (or perhaps magnetic) field, we may conclude that the atoms of the inert gases must be surrounded by very weak fields. If so, there should be little tendency for the atoms of these gases to

combine into molecules or to condense into liquid or solid form; and it is a fact that the inert gases have monatomic molecules and also very low boiling and freezing points.

The properties of the elements adjacent to the inert gases are then accounted for if we suppose, further, that the arrangement of the outer electrons in an atom of an inert gas is an especially *stable* one, *i.e.*, an arrangement of especially low energy. An atom of an element following an inert gas in the table will then contain 1 or 2 extra electrons outside of an inert-gas core (*e.g.*, sodium has 1 and magnesium 2 electrons outside of a neon core); and these extra electrons may well be comparatively easy to detach. The tendency of such atoms to form positive ions would thus be explained. Furthermore, in the solid state these extra "valence" electrons may easily come loose under the attraction of neighboring atoms, functioning, therefore, as free electrons, so that the elements in question ought to be good conductors of electricity, which they are.

An element such as chlorine, on the other hand, could arrange its electrons as they are arranged in argon if it had 1 more electron. Such elements, closely preceding an inert gas in the periodic table, might, therefore, be expected to exhibit a tendency to pick up an extra electron, thereby forming a negative ion. Some of these elements, in fact, form negative ions that are more stable, *i.e.*, have lower energy, than the neutral atom; this is true, for example, of the halogens and of oxygen and sulfur. In the solid state, such atoms would probably contain no electrons with a tendency to become free; thus, the absence of electrical conductivity in these elements would be accounted for.

These ideas furnish a ready explanation for the formation of a compound such as sodium chloride. In combining, the sodium atom loses 1 electron, its remaining 10 electrons then forming the stable configuration that is characteristic of neon (but, doubtless, somewhat more compressed because the nuclear charge of sodium is higher by 1 unit than that of neon). The chlorine atom adds the electron lost by the sodium to its own 17, making 18 electrons arranged in the stable argon configuration (but slightly expanded because of the weaker nuclear charge of chlorine). The electrostatic attraction of the 2 ions thus formed then binds them tightly together into a molecule. When the molecule of sodium chloride thus formed is put into water, the attraction of the ions is weakened and the molecules fall apart into the constituent ions, each with its outer group of electrons in the arrangement characteristic of an inert gas. In a crystal of sodium chloride, also, the grouping into molecules disappears; a crystal of this

type is composed of ions but not of molecules. Thus many chemical and physical facts can be correlated if we make the assumption that the arrangement of the electrons occurring in an inert gas is a peculiarly stable one of low energy.

One of the principal problems of an explanatory theory of the atom is thus to account for the high stability of certain configurations of the electrons. Then an explanation must be found for the occurrence of the long sequences of elements that are observed to be inserted *between* the groups of highly active elements surrounding the inert gases. These intermediate elements do not differ strikingly from each other in chemical properties. Many of the "rare earths" are even difficult to separate chemically. The theory should be such that it leads automatically to all of these relations between chemical properties and atomic number on the basis of as few assumptions as possible.

132. The Static Atom.—While Bohr was endeavoring to develop a theory of atoms on the basis of the older quantum theory, G. N. Lewis and Langmuir suggested a *static* theory of the atom which had some success in correlating chemical and physical facts. It was natural that the physicist should picture an atom as full of *moving* mechanisms, for his information has been acquired from watching atoms *in action* or, at least, from watching the results of their activity, such as the emission of radiation, characteristic or black body, or the scattering of α particles, and the like. The chemist, on the other hand, is somewhat more concerned with the atom at rest. He weighs it in finding what combinations it makes with other atoms; he observes it in crystals apparently at rest except for possible thermal agitation; he thinks of it as occupying a definite position as one of the constituent atoms of various complex molecules, being held in place by certain interatomic forces, perhaps of electrostatic origin. It was difficult to see how a thing so full of whirling mechanisms as the physicist pictured the atom to be could keep as quiet as the chemist found it when he looked at it.

Now there is in the periodic table a remarkable tendency, not yet mentioned, for chemical properties to recur in *groups of eight*. Each inert gas except helium, in fact, forms the terminus of such an octet of elements. The first two octets, lithium to neon and sodium to argon (*cf.* Fig. 93), follow each other immediately; the other three are separated by intermediate sequences of elements. It is reasonable to suppose that each of these groups of eight represents the building up of a particularly stable octet or group of 8 electrons in the outermost shell of the atom. Since we know of no reason why the 8 elec-

trons should not be symmetrically disposed about the nucleus, it was assumed by Lewis and by Langmuir that they were located at the corners of a cube.

The static theory of the atom never got beyond the qualitative stage, however, and it was soon abandoned.

133. Wave Mechanics and the Periodic Table: the First Two Periods.—The key to the periodic table was finally furnished by wave mechanics, especially by means of two new principles that we have seen to be associated with it: the Pauli exclusion principle and the principle of electron spin. These two additions to the theory were not made until 1925. The ideas necessary to understand how the sequence of the elements arises have been described in Sec. 128 above and will be assumed to be familiar. It remains only to add the principle that in an atom in its normal state the electrons are arranged in such a way as to make the energy of the atom as low as possible.

Let us consider the application of these principles to the elements taken in the order of their atomic numbers. The *neutral atom* will always be assumed to be under discussion unless the contrary is stated.

(a) $Z = 1$. *Hydrogen*.—A single electron will have a $1s$ wave function, with $n = 1$, $l = 0$, $\lambda = 0$, $\mu = \pm \frac{1}{2}$; for the energy associated with this wave function is less than that for any wave function with $n > 1$. Degeneracy exists because of the two possible values of μ . This doubling of the 1-electron quantum states due to spin was not considered in our previous treatment of hydrogen. It is of importance, however, when a magnetic field is present.

(b) $Z = 2$. *Helium*.—If 2 electrons are combined with a nucleus for which $Z = 2$, they can both have $1s$ wave functions, with quantum numbers

$$1\ 0\ 0\ \frac{1}{2}, \quad 1\ 0\ 0\ -\frac{1}{2}.$$

The spatial wave function is the same for these two states, being of the form $\psi = Ce^{-2r/a_0}$ where C and a_0 are constants and r denotes distance from the nucleus (cf. Sec. 124). The electronic charge cloud is thus symmetrical about the nucleus in helium. Hence, by Gauss' theorem, there should be no electrical field at all outside of the atom (the total charge on it being zero and symmetrically arranged). Thus, neutral atoms with $Z = 2$ should be chemically inert; they should also exhibit little tendency to associate themselves either in molecules or in a liquid or solid phase. Such atoms should, therefore, form an inert, monatomic gas with very low boiling and melting points. All of these conclusions from wave mechanics agree with the known properties of helium.

In helium, therefore, the K shell is complete. The electronic formula of helium is $1s^2$. All heavier atoms will be expected to contain, next to the nucleus, a complete K shell of this sort. Any additional electrons must then go into outer shells, by the Pauli exclusion principle; for there are no other wave functions with $n = 1$.

(c) $Z = 3$. *Lithium*.—The third electron in lithium, outside of the heliumlike core, must have a wave function with $n = 2$. An approximate calculation shows that the electronic energy associated with a $2s$ wave function is somewhat less than that associated with a $2p$ function for an electron in the atomic central field; hence, we should expect the third electron to have a $2s$ function when the atom is in its normal state. Spectroscopic evidence confirming this conclusion will be described later. The electronic constitution of normal lithium should be, therefore,

$$1s^2 2s,$$

$1s^2$ denoting 2 electrons with $1s$ wave functions. (Since s means $l = 0$, we have $\lambda = 0$, but $\mu = \pm \frac{1}{2}$, so again there is a twofold degeneracy, or a "statistical weight" of 2, in the normal state).

The energy for a $2s$ wave function is found upon calculation to lie much higher than that for a $1s$ function. If the field were a Coulomb field, as for a 1-electron atom, we could use Eq. (186) in Sec. 124, which shows that in such a field the $2s$ state ($n = 2$) lies only a quarter as far below the ionization level as does the $1s$ state ($n = 1$). Hence, the $2s$ electron should be comparatively easy to remove, certainly much easier than either electron in helium. Lithium ought, therefore, to form positive ions easily, and it ought to conduct electricity when in the solid state, the $2s$ electrons of the atoms easily coming loose and functioning as "free" electrons. Both conclusions are in agreement with fact. Since the ion formed by removal of the $1s$ electron is univalent, lithium might be expected to combine chemically with a valence of 1. Such is the fact, as shown by such compounds as Li_2O , LiOH , LiCl .

To remove a second $1s$ electron should require much more energy than to remove the $2s$ or "valence" electron. In harmony with this conclusion, the first ionization potential of lithium is observed to be only 5.38 volts, as against 75.5 volts for the second. Since there is little symmetry to be expected from a single electron in the second shell, the lithium atoms should be surrounded by stray electric fields and should readily group themselves into a condensed phase. The melting point of lithium is 186°C .

The arc spectrum from lithium is predominantly that to be expected from a single electron, as described below (Secs. 142, 147). Evidently

the emission of radiation is chiefly due to the outer or valence electron. The so-called *spark* spectrum, on the other hand, which is ascribed to emission by singly ionized atoms of lithium, contains singlets and triplets of lines like the *arc* spectrum of *neutral helium*. This is what we should expect on the basis of the theory, for a singly ionized lithium atom contains the same number of electrons as neutral helium. The frequencies of corresponding lines are much higher than in helium, however, because the stronger nuclear charge causes all energy levels to lie much lower.

(d) $Z = 4$. *Beryllium*.—Two electrons outside of the K shell can both have $2s$ wave functions but with opposite spins ($\mu = \pm \frac{1}{2}$). The resulting element should be a metal rather like lithium but bivalent, since both $2s$ electrons should come off relatively easily. Such is the case. The oxide, hydroxide, and chloride have the respective formulas BeO , Be(OH)_2 , BeCl_2 .

The second electron is harder to remove than the first, however, since the second helps push the first off by repelling it; the first two ionization potentials of beryllium are 9.3 and 18.1 volts. To remove a third electron, out of the $1s$ shell, requires 153 volts, and to remove the last one, 217 volts.

Again the spectral evidence confirms the theory. The *arc* spectrum of beryllium is a 2-electron spectrum of singlets and triplets like that of neutral helium. In this case both valence electrons play a role in the emission. In the *spark* spectrum, on the other hand, doublet lines like those from neutral lithium are found, emitted by singly ionized beryllium atoms. A few singlet lines are also known in the spark spectrum; it is presumed that they are part of a 2-electron spectrum, emitted by doubly ionized beryllium atoms. Two lines have even been found which are ascribed to triply ionized atoms.

So far it is obvious that the theory succeeds admirably. But perhaps the student will have wondered why beryllium is not, like helium, an inert gas. For the $2s$ wave functions are just as symmetrical as the $1s$ functions. As a matter of fact, even the single $2s$ function for the valence electron in lithium is symmetrical about the nucleus! Only a detailed study of the mathematics of the perturbation theory can really throw light on this question. A satisfactory theoretical answer to such questions can be given, but it is hard to describe in concrete terms.

(e) $Z = 5$. *Boron*.—With 5 electrons in the atom, one of them must have a $2p$ wave function. Boron is a trivalent element, as witness B_2O_3 , B(OH)_3 , BCl_3 . It is not metallic, however; crystals of boron are good insulators. Evidently, with 3 electrons present in the

L shell, conditions are not favorable to the formation of free electrons in the solid state. No simple reason for this is apparent, however.

To remove the $2p$ electron from a boron atom requires only 8.3 volts; to remove the two $2s$ electrons in succession requires 25 and 38 volts, whereas to remove one of the $1s$ electrons as well requires an additional 258 volts.

The next elements in order are most easily understood if we pass them by for the moment and consider neon next.

(f) $Z = 10$. *Neon*.—It is possible to put into an atom two $2s$ electrons (with $\mu = \pm \frac{1}{2}$) and six $2p$ electrons ($l = 1$, $\therefore \lambda = 1$ or 0 or -1 , and with each value of λ , $\mu = \pm \frac{1}{2}$), or 8 in all, but no more electrons with $n = 2$. The L shell is then filled. With every possible value of λ and μ represented, it can be shown that the electronic charge cloud is symmetrical about the nucleus, as it is in helium; and now the symmetry is sufficiently complete so that an inert gas results. Neutral neon in its normal state has the electronic constitution $1s^2 2s^2 2p^6$.

A possible origin thus seems to emerge for all of the octets, or groups of eight elements, pointed out above. The first octet, from lithium to neon, results from the progressive filling of the $2s$ and $2p$ subshells which constitute the L shell. Any outer shell, if filled, will likewise contain an s and a p subshell, with 8 electrons in the two combined. May it be that each octet of elements, ending in a noble gas, represents the filling up of a new pair of s and p subshells?

Laying this question aside for a moment, let us proceed backward from neon to the remaining elements in the first octet.

(g) $Z = 9$. *Fluorine*: $1s^2 2s^2 2p^5$.—If $Z = 9$, the neutral atom contains only 7 electrons in the L shell, or 1 less than enough to fill it. If 1 more electron were added, we should have a negative ion the exterior of which would be a closed shell, as in neutral neon. It cannot be expected that the same loss of energy would occur when an electron is added to a *neutral* atom as when it is added to a *positive ion*; but, on the other hand, the electron is added in the shell for $n = 2$ in an atom with a fairly strong nuclear field ($Z = 9$). We can thus understand the fact that fluorine forms univalent negative ions which are stable, *i.e.*, have lower energy than the neutral atom, and that it exhibits a valence of 1 in chemical combination. There should be little tendency for a *second* extra electron to be bound by the fluorine atom, for it would have to have a wave function with $n = 3$, the energy of which would lie considerably above that for $n = 2$.

Similarly, *oxygen* ($Z = 8$) commonly exhibits a valence of 2 in chemical combination, as in lithium oxide, Li_2O ; and *nitrogen* ($Z = 8$) is commonly *trivalent*, as in lithium nitride, Li_3N . Nitrides are not

actually common, however; more often nitrogen is united with oxygen into a compound negative radical, as in lithium nitrate, LiNO_3 .

The examples of Li_2O and Li_3N illustrate the tendency of elements immediately following a rare gas in the order of atomic number to combine with those just preceding some rare gas. This can be interpreted as arising from a lowering of the energy when closed subshells are formed in both atoms. In forming the molecule Li_2O , for example, an electron passes over from each lithium atom to the oxygen atom, leaving the lithium atoms with closed $1s^2$ shells like helium, while the oxygen acquires a closed $2s^22p^6$ shell, as in neon. The tendency, suggested above in general terms, toward the formation of the electron configuration that is characteristic of an inert gas, is thus interpreted in wave mechanics as a tendency to form closed subshells.

(h) $Z = 6$. Carbon: $1s^22s^22p^2$.—With 4 electrons in the L shell, carbon would seem to have an equal chance either to lose electrons or to add more in an effort to form the closed group of 8. Actually, it probably never does either! The four equal valences of carbon are famous, but its compounds are not formed by a transfer of electrons. Carbon is a typical transition element. In some forms (amorphous carbon, graphite), it exhibits fair metallic conductivity; but in diamond it is an excellent insulator.

At this point, we may with advantage interrupt the discussion of the elements in order to describe in greater detail the wave-mechanical theory of valence.

134. Valence Bonds.—Consider the interaction of 2 atoms as they approach each other from a distance. If we calculate the field produced by either atom at the location of the other, by superposing the field due to the electronic charge upon that due to the nucleus, we shall probably obtain the value zero, since the atoms are electrically neutral. But, if we go a step further in perturbation theory, we obtain the analog of the following classical effect. If the electrons were revolving in classical orbits about the nucleus, their fields would not exactly cancel the field of the nucleus at each instant, although they might do so on the average; there would be a rapidly fluctuating residual field. This field would polarize the other atom and then, reacting on the electric moment thus induced in it, would exert a fluctuating attractive effect upon it. There is an analogous effect in the wave-mechanical theory. The resulting attraction between atoms is responsible for the phenomenon of *cohesion*; it has not, however, a great deal to do with the formation of chemical compounds.

Chemical forces come into play only when 2 atoms approach so closely that their *electronic charge clouds begin to overlap*. New effects then occur. The most important one is due to *electron exchange*,

which was briefly referred to above. The effect of electron exchange is of opposite character according as the 2 electrons involved have the same or opposite spins.

Where the charge clouds due to 2 electrons of the same spin overlap, the exchange effect results in a repulsion between the atoms. For this effect tends to thin out the charge density, piling it up outside the region of overlap. The exclusion principle itself, described in Sec. 128, can be regarded as an extreme case of this effect, 1 electron crowding the other out of the atom completely. Now the region in which the overlap of the charge clouds first occurs will be the region lying between the approaching nuclei and hence will be relatively close to both of them; this region is, therefore, a region of especially low energy for an electron, owing to its attraction by the nuclei. Hence, the exchange effect, by shifting some of the electronic charge cloud away from a region of low energy to a region of higher energy, tends to raise the mutual energy of the atoms. This change in their energy causes the atoms to tend to fly apart, since an isolated system tends to move so as to diminish its energy.

If, however, the overlapping charge clouds belong to electrons of *opposite spin*, the effect of exchange is to draw the charge density *into* the region of overlap at the expense of other regions. To explain how this occurs, we should have to go further into the mathematical theory than is appropriate in this book. Cloud charge is thereby drawn into a region of comparatively low energy, so that the mutual energy of the 2 atoms is lowered. The atoms tend, therefore, to come closer together, as if they attracted each other.

The net effect of exchange will thus depend on the relative arrangement of the electrons in the 2 atoms. As between an electron and a closed subshell, which contains equal numbers of electrons of both spins, the net effect of exchange turns out to be equivalent to repulsion. This is the cause of the observed general *impenetrability* of atoms and molecules; it is due chiefly to repulsion between complete subshells as wholes. If, however, each atom has an electron sticking out on its surface, so to speak, beyond any closed subshells that may also be present, this pair of valence electrons, one in each atom, can adjust their spins so that they are opposite. The exchange effect can then draw the atoms together into a molecule. Atoms thus combined by means of a pair of electrons acting by virtue of exchange are said to be held together in the molecule by a *covalent bond*. Compounds formed by means of covalent bonds are called *homopolar*.

The amount of overlapping of the charge clouds varies greatly according to circumstances. In a collision due to thermal agitation,

a very slight overlapping brings into play sufficient force to cause the molecules to rebound from each other. In the case of argon, for example, the "radius" of the atom as calculated by means of the kinetic theory of gases, which represents about half the distance between 2 atoms in an ordinary collision, is at the value 3.45 of the abscissa in Fig. 92, and thus entirely off the plot. Only a minute part of the charge cloud lies as far as this from the nucleus. In chemical combination, on the other hand, the overlapping may be considerable.

Just as many covalent bonds can be formed for a given atom as there are extra electrons in it outside of closed subshells, so long as these electrons do not more than half-fill the next possible subshell. The atom will thus exhibit chemical valence equal to the number of these extra electrons.

If, however, the extra electrons more than half fill a new subshell, the number available to form covalent bonds tends not to exceed the number of electrons that stand alone in the subshell unmatched by other electrons having the same n, λ, μ but opposite spin. This number is also equal to the number of electrons that would be required to complete the subshell, or to the number of "holes" in it. Thus, when a subshell is more than half full, we can imagine the covalent bonds to be formed by the *holes* in connection with superficial electrons, or perhaps even with similar holes, in other atoms. Electrons in an incomplete outer subshell are called *valence* electrons. Perhaps this term might also be extended to the holes in subshells that are more than half full.

Covalent bonding is not, however, the only type of chemical bonding. If the charge cloud of an extra electron lying outside of a closed subshell on one atom comes opposite a hole in an incomplete subshell in another atom, the charge cloud may be drawn into the hole, so to speak, thereby getting closer to the nucleus of the second atom than it can to its own nucleus and therefore moving into a location of lower energy. In this way 2 ions are formed, one charged positively and the other negatively. The 2 ions will then be drawn together into a molecule by electrostatic attraction. In this way are explained the *ionic* compounds between strongly electropositive and strongly electronegative elements, examples of which have been mentioned above.

Both modes of formation result in the same number of valence bonds per atom.

Furthermore, in reality all grades of transition occur between the two types; or, rather, the mathematical treatment is best described

by saying that all actual chemical bonds are partly of one type and partly of the other. Bonds between strongly electropositive and strongly electronegative elements (*i.e.*, between those closely following or closely preceding an inert gas) are predominantly ionic (except in the case of hydrogen!). In general, however, most chemical bonds are predominantly covalent. Molecules such as O_2 and N_2 are also held together by predominantly covalent bonds.

The subject of molecular combination and the related topic of the formation of crystals are complicated subjects and the student is referred for further information in regard to them to other books.¹

135. The Third Period of the Periodic Table.—The second octet of elements, from $Z = 11$ to $Z = 18$, parallels so closely the first octet that brief comments will suffice. (Corresponding elements are exhibited clearly in Fig. 93.) *Sodium* ($Z = 11$, $1s^2 2s^2 2p^6 3s$) contains a single valence electron outside the neon core, but this electron has a $3s$ wave function, as against the $2s$ function for the valence electron of lithium. Sodium is univalent, as in $NaOH$ and the familiar $NaCl$. Its arc spectrum is a typical 1-electron spectrum. In general, it resembles lithium closely. *Magnesium* ($Z = 12$) is a bivalent element similar to beryllium. It burns with a brilliant white flame to form the oxide MgO . It has an arc spectrum of singlets and triplets resembling that of helium. *Aluminum* ($Z = 13$) is trivalent, like boron, but it is metallic and an excellent conductor of electricity. The "sesquioxide," Al_2O_3 , occurs in crystalline form as corundum and sapphire and the ruby. *Silicon* ($Z = 14$) is a good deal like carbon. The dioxide, SiO_2 , however, which occurs in quartz, is a substance of extremely high melting point, whereas the analogous compound, CO_2 , is a gas! Only an elaborately refined application of wave mechanics can explain contrasts such as these. *Phosphorus* ($Z = 15$), although chemically much more active, forms compounds analogous to those of nitrogen. In the poisonous and bad-smelling gas, phosphine or PH_3 , phosphorus is trivalent, just as nitrogen is in gaseous ammonia, NH_3 . *Sulfur* ($Z = 16$) corresponds to oxygen. In H_2S , it is bivalent, just as oxygen is in H_2O . *Chlorine* ($Z = 17$) is univalent and easily forms negative ions in solution; in general, it resembles fluorine closely but is less active. Finally, in *argon* ($Z = 18$), we reach again an atom composed of closed subshells, with electronic formula $1s^2 2s^2 2p^6 3s^2 3p^6$. Thus, argon has complete K and L shells and, outside of these, two more complete subshells. The next or M shell is not yet complete, since there are no $3d$ electrons.

¹ Cf. PAULING, L., "The Nature of the Chemical Bond," 1940; SLATER, J. C., "Introduction to Chemical Physics," 1939.

The symmetry of the $3p^6$ configuration, however, is evidently enough to make the external field around argon very weak, so that it behaves as an inert gas with a very low boiling point ($-186^\circ\text{C}.$).

136. The Fourth and Fifth Periods.—At $Z = 19$, we should naturally expect the addition of an electron in a $3d$ state (*i.e.*, $n = 3$, $l = 2$). For, in the 1-electron atom, we found the energy to increase regularly with increasing values of n ; hence, we might anticipate that even in our modified field any wave function with $n = 4$ would correspond to a higher electronic energy than a $3d$ function. But the next element in order, potassium, closely resembles sodium and lithium, not only in chemical properties but also in its spectra even to the finest details. There is abundant reason to believe that atomic spectra in the visible and ultraviolet regions are emitted by electrons in the periphery of the atom. These spectra, therefore, furnish valuable information concerning the state of the outermost electrons. Can it be, then, that the valence electron of potassium is not a $3d$ but a $4s$ electron?

The conclusion that it is a $4s$ electron was drawn from the empirical facts before wave mechanics came on the scene; and it has been confirmed theoretically by several approximate calculations. Wave functions for $l = 0$ and $l = 2$ are very dissimilar. The radial wave functions for electrons in the modified central field resemble qualitatively those for the 1-electron atom, a few of which are given in Sec. 124 and are plotted (partially) in Fig. 89. The $3d$ radial function, like any function for $l = n - 1$, shows a single maximum. The $3s$ function, on the other hand, is much more drawn out, corresponding to the Bohr-Sommerfeld elliptical orbit of maximum eccentricity; it has 3 (*i.e.*, $n - l$) maxima and, like all the s functions, is not zero at the nucleus. The $4s$ function is similar to the $3s$ function but has 4 maxima. One can well believe that the energies associated with such different functions may be quite different, although it is not easy to see which energy should be higher, or, indeed, that there actually should be a difference (and in the Coulomb field, of course, there is not!). Approximate calculations show that in *potassium* ($Z = 19$) the theoretical energy of a $4s$ electron actually is below that of a $3d$ electron; hence, in the normal state of the potassium atom, the last electron should have a $4s$ wave function. The formula for potassium is thus $1s^2 2s^2 2p^6 3s^2 3p^4 s$, and the facts are fully accounted for.

The next element, *calcium* ($Z = 20$), has two $4s$ electrons and corresponds rather well to magnesium in the preceding octet. Thus, we appear here to be witnessing the building up of another octet. The next four elements, *scandium* (21), *titanium* (22), *vanadium* (23), and

chromium (24) show some degree of resemblance to the corresponding elements in the preceding period (Al, Si, P, S). The next element after these, however, *manganese* (25), is not at all like chlorine [cf. $\text{Mn}(\text{OH})_2$!]; and, in the group that follows manganese, consisting of *iron* (26), *cobalt* (27), and *nickel* (28), we encounter something definitely new.

The accepted theoretical explanation of this failure to finish out an octet at this point in the periodic table is the following. As far back as scandium ($Z = 21$), there is theoretical evidence for the belief that the electron added last goes into the $3d$ subshell; otherwise, it would have a $4p$ wave function, the $4s$ subshell being full, and the energy of a $4p$ electron lies above that of a $3d$. Spectral evidence confirms this theoretical conclusion. The next step in atom-building after calcium ($Z = 20$) is, therefore, the addition of ten $3d$ electrons in succession. Only in chromium and copper are there irregularities; in these atoms one of the $4s$ electrons is replaced by an extra $3d$.

Copper ($Z = 29$, $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s$) is the first element that contains, in its normal state, a complete $3d$ subshell. It thus has the *K*, *L*, and *M* shells all complete, and 1 valence electron outside of these. The last $3d$ electron, however, added out of turn as if in haste to complete the $3d$ subshell, is rather lightly bound and comes off rather easily, so that copper frequently exhibits a valence of 2 as an alternative to its expected valence of 1 (cf. cuprous oxide, Cu_2O , and cupric oxide, CuO).

In *zinc* ($Z = 30$, $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2$), the $4s$ subshell is completed. Thus zinc, like calcium and magnesium, has 2 valence electrons outside of closed subshells. These three elements are much alike both chemically and spectroscopically. In magnesium and zinc, however, the core of the atom consists of entire closed *shells* (*K* and *L* in magnesium, *K*, *L*, and *M* in zinc); perhaps this is the reason why the chemical resemblance is, in general, closest between these two elements. However, whereas the oxides MgO and CaO react with water to form the hydroxide, $\text{Mg}(\text{OH})_2$ or $\text{Ca}(\text{OH})_2$, respectively, ZnO occurs naturally in a zinc ore. In this respect, magnesium resembles calcium more closely. The carbonates all occur naturally as minerals, MgCO_3 as magnesite, CaCO_3 as limestone, ZnCO_3 as smithsonite.

Beyond zinc, the formation of an octet sequence is definitely resumed, the $4p$ subshell being progressively filled. *Gallium*, *germanium*, *arsenic*, *selenium*, and *bromine* are good chemical analogs of aluminum, silicon, phosphorus, sulfur, and chlorine. Finally, at $Z = 36$, we have the next inert gas, *krypton*, with electronic formula $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6$. Thus for the third time an inert gas

occurs just when a new p subshell has been completed, in accord with the rule that we guessed in connection with the study of neon.

The *fifth period* of 18 elements, from $Z = 37$ to $Z = 54$, corresponds so closely to the fourth that detailed comment is unnecessary. Two $5s$, ten $4d$, and six $5p$ electrons are added, the final element in the period being the inert gas xenon. The analogs in the fifth period of iron, cobalt, and nickel in the fourth period are the three elements ruthenium, rhodium, and palladium. Although the latter elements are not ferromagnetic, they have many chemical similarities to the ferromagnetic elements Fe, Co, Ni. The explanation of ferromagnetism on the basis of wave mechanics is intricate; that it occurs at all is rather of the nature of an accident. Silver, the analog of copper in the preceding period, is univalent, not exhibiting the abnormal alternative bivalency.

137. The Periodic Table Concluded.—*Cesium* (55) and *barium* (56) start the sixth period with one and two $6s$ electrons, respectively, outside of an argon core; they constitute the fifth pair, consisting of an alkali and an alkaline earth, to be formed by the addition of two s electrons in succession (the first four pairs being lithium and beryllium, sodium and magnesium, potassium and calcium, rubidium and strontium).

Lanthanum (57) then follows the precedent set by scandium and yttrium in the preceding periods and adds a $5d$ electron. But beyond this point something new occurs again. With $n = 4$, $4f$ states with $l = 3$ are possible. From *cerium* (58) to *lutecium* (71), the $4f$ electronic energy probably lies below the $5d$; hence, over this range in Z , fourteen $4f$ electrons are added in succession

$$[14 = 2 \times (2l + 1) = 2 \times (6 + 1)].$$

These electrons lie rather deep in the atom, however; their wave functions have a single maximum, whereas the $6s$ and $5d$ functions are considerably drawn out and thus project more on the exterior. Hence the addition of the $4f$ electrons causes very little change in chemical properties, and we have the group of the chemically similar, and in part almost indistinguishable, *rare earths*.

Only when the $4f$ subshell has been completed is the filling of the $5d$ subshell resumed. From *hafnium* (72) to *mercury* (80), the remaining nine $5d$ electrons are added, with a few irregularities. In *gold* ($Z = 79$), the $5d$ subshell is completed out of turn, in analogy with copper and silver. Gold, however, manages to combine as a trivalent element as well as a univalent one (e.g., AuCl_3 as well as AuCl). The next element, *mercury* ($Z = 80$), is analogous to beryllium, magnesium, zinc, and cadmium. Then the $6p$ electrons are added, completing an

octet of elements (except for the missing 85) which corresponds weakly to previous octets and ends with the last known rare gas, *radon* or radium emanation ($Z = 86$, $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 4f^{14} 5s^2 5p^6 5d^{10} 6s^2 6p^6$).

Beyond radon, we find a fragment of a *seventh* period, including radium (88), thorium (90), and uranium (92). All of the known elements in this period, however, are radioactive and therefore unstable. It would seem that wave mechanics, as applied to the electrons, imposes no obstacle to the occurrence of elements beyond uranium; the electrons would "know how" to group themselves about a nucleus with $Z > 92$, for example; but the appropriate nuclei encounter some difficulty in existing (at least for any length of time).

From this discussion it is evident that wave mechanics, as applied to the electrons in atoms of varying atomic number, leads without the

TABLE III.—COMPLETED ELECTRON SHELLS AND SUBSHELLS

n, l	X-ray symbol	Electron symbol for subgroup	Electrons in subgroup $[2(2l + 1)]$	Element at which subgroup is completed	Electrons in completed n group, $2n^2$	Element at which n group is completed
1, 0	K	$1s$	2	He(2)	2	He(2)
2, 0	L_I	$2s$	2	Be(4)	8	Ne(10)
2, 1	$L_{II,III}$	$2p$	6	Ne(10)		
3, 0	M_I	$3s$	2	Mg(12)	18	Cu(29)
3, 1	$M_{II,III}$	$3p$	6	Ar(18)		
3, 2	$M_{IV,V}$	$3d$	10	Cu(29)		
4, 0	N_I	$4s$	2	Zn(30)	32	Lu(71)
4, 1	$N_{II,III}$	$4p$	6	Kr(36)		
4, 2	$N_{IV,V}$	$4d$	10	Pd(46)		
4, 3	$N_{VI,VII}$	$4f$	14	Lu(71)		
5, 0	O_I	$5s$	2	Cd(48)	50	
5, 1	$O_{II,III}$	$5p$	6	Xe(54)		
5, 2	$O_{IV,V}$	$5d$	10	Au(79)		
5, 3	$O_{VI,VII}$		14			
5, 4	$O_{VIII,IX}$		18			
6, 0	P_I	$6s$	2	Ba(56)		
6, 1	$P_{II,III}$	$6p$	6	Rn(86)		
6, 2	$P_{IV,V}$		10			
7, 0	Q_I	$7s$	2	Ra(88)		

addition of special assumptions to a sequence of chemical properties such as is actually observed.

The probable electronic constitution of all the elements, when in their normal state, is shown for reference in Appendix III. An outline showing the completed shells and subshells is given in Table III. For convenience in future reference, the table includes also the designations of the subshells that are used in the discussion of X-rays.

OPTICAL SPECTRA

In previous chapters we have become acquainted with a few features of atomic spectra. We saw in Chap. VI that a general characteristic of such spectra is the occurrence of series of lines converging toward short-wave-length limits. In Chap. VII, we have seen how wave mechanics was able to supply a quantitative theory of the general features of spectra emitted by 1 electron in the field of a nucleus, yielding the same formulas as did the original quantum theory of Bohr. A brief but more systematic account will now be given of some representative phenomena of the spectra emitted by atoms or molecules containing more than 1 electron, and of their theoretical interpretation by means of wave mechanics. In connection with this discussion, we shall also return briefly to the 1-electron atom in order to discuss the "fine structure" of its lines, which is due to spin and other relativistic effects. In the present chapter, however, the discussion will be limited to spectra whose emission is connected with the motion of the exterior electrons of atoms or molecules or with nuclear motions; these spectra occur in the infra-red, visible, or ultraviolet regions. The high-frequency radiation known as X-rays will be reserved for special treatment in a later chapter.

138. Atomic and Molecular Spectra.—The variety of spectral lines that can be emitted by an atom containing several or many electrons, although very large, is less than might be anticipated, for several reasons.

For one thing, in discussing atomic spectra, it suffices to discuss the spectra emitted by *neutral* atoms. For it is a universal rule that *the same types of spectra that are emitted by neutral atoms of given atomic number Z are emitted also by singly ionized atoms of atomic number $Z + 1$, by doubly ionized atoms of number $Z + 2$, by triply ionized atoms of number $Z + 3$, and so on.* The principal difference is that the frequencies of corresponding lines increase with each stage of ionization, owing to the stronger field in which the active electron moves. The physical basis for this rule lies in the fact that the atoms

specified all have the same number of electrons, and these electrons arrange themselves in the same types of configurations. Such atoms are called *isoelectronic*. The similarity of their spectra simplifies the theoretical discussion; but it also leads to great complications in the observed spectrum as emitted by a given element when excited in a high-potential discharge, since in such a case the different spectra emitted by atoms in various stages of ionization are all jumbled together.

The theoretical basis for the treatment of many-electron atoms was sketched in Secs. 127 to 129. As is there explained, the first step is to construct a simplified approximate theory. In this theory, the electrical field due to the nucleus is replaced by another central field which includes the average effect of the fields of the remaining electrons, the latter being supposed for this purpose to be replaced by an equivalent charge cloud. Then, subsequently, corrections for the finer effects of the electrostatic interaction of the electrons are introduced by the methods of perturbation theory. In the first stage of the treatment, each electron is assumed to move in the central field, as if the others were not there. When the atom is in a stationary state, therefore, each electron can be supposed to be represented by a certain 1-electron wave function.

In treating the radiation from atoms, however, it is not necessary to keep track of *all* of the electrons. The wave-mechanical theory of radiation leads to a simplifying principle which has the following two aspects: (1) *in radiation phenomena, the various electrons act independently of each other*; and (2) *in so far as it is possible to regard the electrons as represented by a set of 1-electron wave functions, jumps occur only between those stationary states of the atom which differ in just one of the electronic wave functions*. The two statements are often condensed into the single statement that, as a rule, only 1 electron jumps at a time.

The complexity of atomic spectra is greatly reduced by this double-headed principle. If the atom has 1 or more valence electrons outside of closed subshells, the electrons in the subshells play no part in most of the radiation phenomena with which we shall be concerned; usually only the valence electrons are active. Furthermore, only 1 of these valence electrons jumps at a time (as a rule). The valence electrons that do not jump, if there are such, cannot be ignored entirely, however; as we shall see, their presence has a great deal of influence upon the characteristics of the atomic energy levels between which the jumps occur. If there are no valence electrons, as in an inert gas, then for the emission of spectral lines it is necessary

that an electron be first transferred by some process of excitation from a closed subshell—from the outermost subshell, in the cases considered here—to a 1-electron wave function of higher energy.

The complexity of many-electron spectra is further reduced by what are known as *selection rules*. Wave mechanics not only tells us what frequencies can occur in the spectrum but it also furnishes a certain amount of information concerning the relative intensities of the various spectral lines. It gives us not only the energies of the various atomic states but also information concerning the relative probability of jumps between them. Sometimes the probability of a given jump is excessively small, or even zero; in that case, the corresponding line is not seen in the spectrum. A rule stating that jumps between a certain pair of levels do not occur is called a selection rule.

The principal selection rules are closely associated with the angular momentum of the electrons in the atom. The quantum states of complex atoms, like those of 1-electron atoms, are characterized by special features in regard to angular momentum. It will facilitate understanding of the subject if we first discuss the treatment of angular momentum in wave mechanics. The results obtained from the theory are easily visualized; and we shall go no further than is necessary for our purpose.

139. Angular Momentum in Wave Mechanics.—When several electrons are present in the same atom, we should not expect that the angular momentum of each individual electron will be constant, because of the forces on each one due to the repulsion of the others. Classical analogy, however, does lead us to expect that the *resultant* angular momentum of *all* of the electrons about the nucleus, including the angular momentum due to spin, will have a constant value; for, in classical mechanics, the angular momentum of any isolated system is conserved.

This expectation is confirmed by the wave-mechanical theory, but only within characteristic limitations. In the new theory, the various components of any angular momentum are mutually incompatible quantities, so that a form of the indeterminacy principle applies to them. The component of an angular momentum in any chosen direction is susceptible of measurement¹; the result will be a quantized value,

$$\frac{Mh}{2\pi},$$

¹ For a method of measuring one component, by means of the associated magnetic moment, see the description of the Stern-Gerlach experiment in Sec. 162.

where M is an integer or half-integer, positive or negative, or zero. (By "half-integer" is meant half of some integer, such as $\frac{1}{2}$, $-\frac{3}{2}$.) It is impossible, however, according to the theory, to measure with precision *two* components of an angular momentum at the same time. Hence *no direction* can be assigned to the angular momentum of an atomic system as a vector. After one component, say G_x , has been measured, the system is left in such a state that all perpendicular components, such as G_y or G_z , are indefinite, with an average expectation of zero. That is, if a subsequent measurement of a perpendicular component, say G_y , were made, any one of several values might be obtained; and, if the whole experiment were repeated many times, both G_x and G_y being measured in succession each time and the same value being obtained each time for G_x , the values obtained for G_y would be sometimes positive and sometimes negative, their average coming out zero. There is just one exceptional case, that in which all three components are definitely zero. These features stand in strong contrast to the classical picture.

It should be possible, however, according to the theory, to measure at the same time the *component* of an angular momentum in *any chosen direction*, say G_x , and also the *sum of the squares of all three components*, or

$$G_x^2 + G_y^2 + G_z^2.$$

The latter is ordinarily regarded as the square of the total magnitude of the momentum and is denoted by G^2 . From the values thus found for G_x and G^2 we can then calculate a value for the square of the component perpendicular to our chosen direction, or $G_y^2 + G_z^2 = G^2 - G_x^2$.

140. Total Angular Momentum and Its Selection Rules.—When an atom is in a definite quantum state, wave mechanics leads to the following conclusions concerning its angular momentum. *As long as no external forces act on an atom, its quantum states can be so chosen that, when the atom is in one of them, its component of total angular momentum (including that due to spin) about any chosen axis is*

$$M \frac{h}{2\pi}$$

and the magnitude of its total angular momentum, which is defined as $(G_x^2 + G_y^2 + G_z^2)^{1/2}$, is at the same time

$$\sqrt{J(J+1)} \frac{h}{2\pi},$$

where J and M are two quantum numbers characteristic of the quantum state. The number J is a positive integer or half-integer or zero, and M

has one of the $2J + 1$ integrally spaced values between $M = -J$ and $M = J$, inclusively. Thus, M is either integral or half-integral according as J is. For example, if $J = 0$, $M = 0$; if $J = \frac{1}{2}$, $M = \frac{1}{2}$ or $-\frac{1}{2}$; if $J = 1$, $M = 1$ or 0 or -1 ; etc. Often M_J is written for M .

These propositions represent the utmost that can be stated by way of analogy with classical theory. The relationship of the quantum numbers J and M can be visualized, if desired, with the help of a *vector diagram* [cf. the similar treatment of orbital momentum in Sec. 124(b)]. A vector of length J can be drawn at such an angle that its projection on the chosen axis is M (Fig. 94). For $M = \pm J$ the vector will lie along the axis.

The J -vector so drawn can be thought of as representing, in a sense, the total angular momentum of the atom G_J . To visualize the fact that G_J has no definite direction in wave mechanics, we may imagine the whole figure to whirl at arbitrary speed about the axis, so that all components of the J -vector perpendicular to the axis are as often negative as positive and average to zero. To represent all of the wave-mechanical values

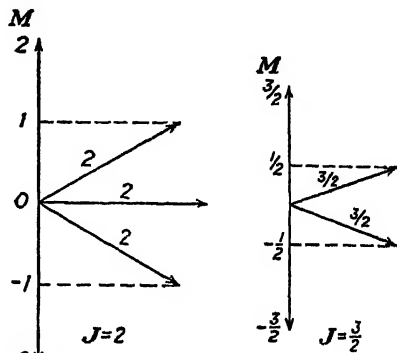


FIG. 94.

completely, the vector representing G_J should be drawn of length $\sqrt{J(J+1)}$; or it may itself be drawn of length $\sqrt{J(J+1)} (h/2\pi)$, and its projection of length $M(h/2\pi)$. When so drawn, the vector is necessarily somewhat inclined to the axis even when $M = \pm J$. The square of its component perpendicular to the axis then represents correctly the value of the square of the corresponding component of the angular momentum. For use in thinking about the quantum states and their quantum numbers, however, the simpler J -vector of magnitude J (or $Jh/2\pi$) is more convenient.

The method of representing the relation between J and M just described constitutes a natural generalization of that used for the orbital angular momentum of a single electron in Sec. 124(b).

Selection Rules.—Wave mechanics leads to the following selection rules for J and M , which are of great importance in spectroscopy. In any atomic jump associated with the emission or absorption of a photon, J may remain unchanged or it may increase or decrease by unity, but not by any larger amount. That is, if J_1 and J_2 are the values of J for the two states between which the jump occurs, then

either $J_1 = J_2$ or $J_1 - J_2 = 1$ or $J_1 - J_2 = -1$. No jump can occur, moreover, from a state with $J = 0$ to another state with $J = 0$. The first rule stated holds also for M . In symbols:

$$\Delta J = 0 \text{ or } \pm 1 \text{ (and not } 0 \rightarrow 0\text{)}; \quad (191a)$$

$$\Delta M = 0 \text{ or } \pm 1. \quad (191b)$$

These selection rules are not always strictly obeyed, however; sometimes jumps occur for which $\Delta J = \pm 2$ or $\Delta M = \pm 2$. Spectral lines due to such jumps are mostly very weak.

Even these selection rules find an analogy in classical theory (and were first proposed on the basis of this analogy), at least if spin is omitted. Orbital angular momentum, in classical mechanics, is associated with rotation. Suppose that a body capable of rotation contains an electrical vibrator of some sort; and let this vibrator emit electromagnetic radiation of frequency ν when the body is at rest. Then, if the body is set rotating with angular velocity ω , the radiation emitted will contain the three frequencies ν , $\nu + \omega$, $\nu - \omega$. The reason is similar to that for the analogous effect of a magnetic field upon a classical vibrating electron, which was described in Sec. 43 in connection with the classical explanation of the Zeeman effect. The three classical frequencies $\nu + \omega$, ν , $\nu - \omega$ correspond (in emission) to the three possible changes in M . This is an example of Bohr's correspondence principle.

141. Orbital and Spin Angular Momenta.—Before taking up the application of these ideas to spectra, it is necessary to say something concerning the angular momentum due to spin and that due to the wave-mechanical equivalent of orbital motion, taken separately. As was explained in Sec. 124, angular momentum can be treated as a vector quantity; hence, we can regard the total resultant angular momentum as the vector sum of the orbital and spin momenta of the electrons. We could form this sum by first adding up the orbital momenta into a *resultant orbital angular momentum*, which we shall denote by G_L , then adding the spin momenta vectorially into a *resultant spin momentum*, G_S , and then finally adding these two into the grand resultant, G_J :

$$G_J = G_L + G_S.$$

If the spin magnetic moment and the orbital motion influence each other, then, according to classical theory, only their resultant G_J remains constant, but not G_L^2 and G_S separately. An interesting analogy is presented by the earth in its combined motion of rotation on its axis (analogous to spin) and of revolution in its orbit about the sun. The differential attraction of the sun on the equatorial bulge

of the earth causes its axis to swing off sideways, as if the earth were a huge gyroscope, the resulting slow change in direction of the earth's axis giving rise to what is called the precession of the equinoxes. The *magnitude* of the earth's angular velocity of rotation, or of the associated angular momentum, is not changed by this effect; but, if we draw almost any fixed axis and watch the component of the earth's angular momentum about this axis, this component will be observed to change. At the same time a compensating change occurs in the corresponding component of the angular momentum which the earth possesses because of its orbital motion, although the total amount of this angular momentum too remains constant; the plane of the orbit merely undergoes a slight tipping. The vector sum of the rotational and orbital angular momenta of the earth, however, remains constant. If it were not for the gravitational effect in question, the rotational and orbital momenta taken separately would also be constant.

A partially similar picture is presented by the conclusions drawn from wave mechanics in regard to the angular momentum of electrons in atoms. In general, only the *total* angular momentum is fixed when the atom is in one of its quantum states. The cause of this restriction lies in the spin-orbit terms in the wave equation, which may be regarded as representing an interaction between the magnetic moment that is associated with spin and the orbital motion. In many of the lighter atoms, however, the spin-orbit effect is comparatively small. Then we have an approximation to the ideal case in which there is no spin-orbit interaction at all. In this latter case two further, very important methods of classifying atomic energy levels by means of angular momentum become available.

In the first place, *in the absence of spin-orbit interaction*, the total *orbital* angular momentum of all of the electrons taken together can be fixed or quantized. For its quantum numbers, we may write L and Λ (more commonly, M_L is used in place of Λ); the magnitude of the resultant orbital angular momentum is then $\sqrt{L(L+1)}\hbar/2\pi$, and its component along the chosen axis is $\Lambda\hbar/2\pi$. The number L is a positive integer or zero, and the $(2L+1)$ possible values of Λ are integrally spaced from L to $-L$. Thus, if $L=0$, $\Lambda=0$; if $L=1$, $\Lambda=1, 0$, or -1 . The quantum number Λ is not of much practical interest, however. In spectroscopic work, it is customary to indicate the value of L by means of letters. Just as for 1 electron the values $l=0, 1, 2, \dots$ are indicated by s, p, d, \dots (Sec. 128), so the values $L=0, 1, 2, \dots$ are indicated by S, P, D, \dots :

$$\begin{array}{cccccccc} L & = & 0, & 1, & 2, & 3, & 4, & 5, & \dots \\ & & S & P & D & F & G & H & \dots \end{array}$$

the atomic states. To keep our notation uniform, however, it may be preferable to use the letter L , understanding that here $L = l$. The corresponding capital letters will be used for the terms. The letter representing an energy level may be preceded by a number indicating the value of the principal quantum number n for the valence electron.

The array of possible LS levels, arising from changes in the electronic function of the valence electron, is thus as shown (in part) in Fig. 95. The S terms ($L = l = 0$) begin with $n = 1$, the P terms ($L = l = 1$) with $n = 2$, and so on; for always $n > l$. The allowed transitions between these levels are indicated in the figure by oblique

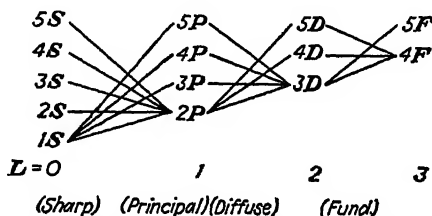


FIG. 95.—Partial LS term system of a typical alkali metal (in actual cases some of the lowest terms shown are missing). Lines indicate allowed transitions.

lines. According to the selection rule for L , as stated above, ΔL cannot exceed unity; hence, no jumps can occur, for example, between S and D levels. In the case of a single active electron, furthermore, the restriction is actually a little more stringent; for, in any case, it is a special rule that the l of the jumping electron must change by unity. Hence, when $L = l$, we must have

$$\Delta L = \pm 1,$$

jumps for which $\Delta L = 0$ being excluded. Thus, jumps cannot occur in such spectra between two S levels or between two P levels, and so on.

For the quantum number S , we have here $S = s = \frac{1}{2}$ (Sec. 126). We do not need to bother with S , however, since it has only one value and is constant.

The scheme just described should obviously be applicable also to the energy levels of the hydrogen atom, which likewise contains a single valence electron. In our previous discussion of hydrogen, we ignored the very slight variation of the energy with l , or L , due to spin and relativistic effects. Hence all S, P, D, \dots levels corresponding to a given value of n fell on top of each other, and no distinction between them needed to be made. We shall return to this point later (Sec. 148).

The theoretical array of energy levels just indicated is strongly reminiscent of the spectral terms employed by Rydberg to represent

the chief series in the spectra of the alkalis. According to Eqs. (128a,b,c) in Sec. 97, the wave numbers of the lines of these series are given by the following formulas:

$$\text{Sharp:} \quad \bar{\nu} = \frac{R}{(1+P)^2} - \frac{R}{(m+S)^2}, \quad m \geq 2;$$

$$\text{Principal:} \quad \bar{\nu} = \frac{R}{(1+S)^2} - \frac{R}{(m+P)^2}, \quad m \geq 1;$$

$$\text{Diffuse:} \quad \bar{\nu} = \frac{R}{(1+P)^2} - \frac{R}{(m+D)^2}, \quad m \geq 2.$$

Thus, the wave numbers are the differences between expressions or "terms" of just three different types:

$$\frac{R}{(m+S)^2}, \quad \frac{R}{(m+P)^2}, \quad \frac{R}{(m+D)^2},$$

where R is the Rydberg constant (Sec. 96). By varying m in these expressions we obtain three sets of terms similar to the three sets of energy values as derived from wave mechanics. It is reasonable to assume that Rydberg's terms correspond to these latter sets of energy values. The Rydberg terms converge to 0 as $m \rightarrow \infty$, just as the theoretical energy levels converge to zero as $n \rightarrow \infty$, provided we agree to call the energy 0 at the ionization limit. The terms are positive, however, whereas the energies are all negative. Hence, the terms must represent the *numerical values* of the energy without the negative sign. As has been said before, positive numbers obtained in this way are commonly used by spectroscopists, under the name of spectral terms, to represent energy levels. In the formulas for $\bar{\nu}$, then, the *second* Rydberg term refers to the *initial* atomic state for the transition in question; the second term together with its minus sign equals the energy of the initial state. The *first* term in the formula refers to the *final* state; its (negative) energy is subtracted from the energy of the initial state to obtain $h\nu$ for the transition.

The sequence of terms represented by the first of the three expressions written above, $R/(m+S)^2$, represents, therefore, the sequence of initial states for the various lines of the sharp series. These terms are observed to combine only with terms in the second sequence, $R/(m+P)^2$. Hence, the first of these two sequences of terms must be those derived theoretically with $L = 0$, or the " S " terms; for the latter is the only theoretical sequence of terms that combines with no other sequence. The second of the Rydberg sequences must then be the P terms, with $L = 1$; for S terms combine only with P terms, by the selection rule for L . The *sharp* series of lines thus arises from

transitions between an S level (exclusive of the lowest) and the lowest of the P levels. The *principal* series, on the other hand, arises from transitions between various P levels and the lowest S level. The lowest S level is the lowest possible value for the energy and represents the normal state of the atom.

The third set of levels, represented by $R/(m + D)^2$, which combines with P but not with S levels, must, then, be the D levels ($L = 2$). The *diffuse* series of lines thus arises from transitions between the various D levels and the lowest P level. A fourth observed series of lines, named long ago the "fundamental" series, has a convergence limit equal to the difference between the convergence limit of the diffuse series and the frequency of the first line of this latter series, *i.e.*, to the difference in energy between the state of ionization and the lowest D level. The final state for the corresponding transitions must, therefore, be the lowest D level. But the initial energies for the fundamental series are not those of the P terms. They must be, then, the F levels, with $L = 3$. The fundamental series is thus due to transitions from an F level to the lowest D level.

In this way, we arrive at the following interpretation of the chief series of the alkali metals:

Sharp,	$nS \rightarrow n_0P,$	$n > n_0;$
Principal,	$nP \rightarrow n_0S;$	
Diffuse,	$nD \rightarrow n_0P;$	
Fundamental,	$nF \rightarrow n_0D.$	

Here the lines of each series are obtained by assigning various values to $n; n_0$ before any letter denotes the lowest value of n that can occur with that letter in the kind of atom under consideration.

The identification thus made of the terms in the alkali spectra constitutes the historical reason for the use by spectroscopists of the mysterious letters S, P, D, F (from the words "sharp, principal, diffuse, fundamental") to represent $L = 0, 1, 2$, or 3 (or the corresponding small letters for 1 electron). For higher values of L it was agreed later to continue down the alphabet, skipping J .

As an example, a diagram of the known energy levels for neutral sodium is shown in Fig. 96. Each level is represented in this diagram by a short horizontal line, the levels being grouped into sequences according to the value of L , in contrast to the arrangement in Fig. 70(b). The diagram shows also the splitting of the P levels, which is discussed in Secs. 144 to 147 below; the subscripts on the letters, denoting J values, refer to this splitting. As usual, the energies are measured downward from zero at the state of bare ionization (with

no kinetic energy in the removed electron) and are expressed in wave-number units. The numbers written opposite the levels are the accepted values of n , which differ somewhat, for reasons explained in the next section, from Rydberg's original number m . At the right are shown for comparison some of the levels of hydrogen, the lowest being far below the sodium levels and therefore not shown.

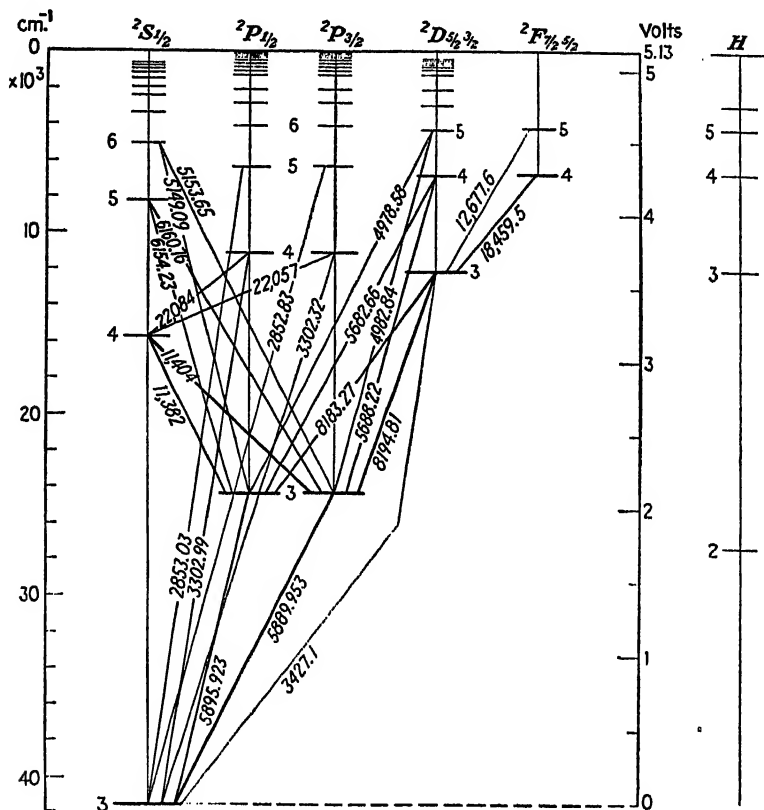


FIG. 96.—Energy-level diagram for neutral sodium, showing also the origin of the principal spectral lines. Some of the levels of hydrogen are shown for comparison. (Wave lengths are in angstroms. The numbers opposite the levels are values of n —see Sec. 143.)

The most important spectral lines are also indicated on the diagram by oblique lines drawn between the appropriate energy levels, with wave lengths attached; the widths of these lines in the diagram give a very rough idea of the relative intensities of the spectral lines as commonly observed. One line is shown that violates the ordinary selection rule for L ($3D \rightarrow 3S, \Delta L = 2$).

The values of some of the terms, expressed in cm.^{-1} , are also given in Table VI (page 363).

143. The Term Energies of the Alkali Metals.—If the mathematical difficulties could be overcome, it should be possible to calculate the energy levels from the wave equation. Unfortunately, however (or fortunately), this cannot be done with any degree of precision, so that experiment is still necessary. Considerable interest attaches nevertheless to approximate theoretical values of the term energies. Such values can be obtained, for a neutral alkali atom, by disregarding the spin-orbit effect and solving the wave equation for the valence electron moving in a suitable central field.

As described in Sec. 127, the field in which the valence electron is assumed to move is chosen so as to allow for the average effect of the remaining electrons in the atom. At points well outside the atom, this field must approximate closely to the field of a proton. For, by ordinary electrostatic reasoning, based upon Gauss' law, it can be shown that a spherically symmetrical distribution of charge produces the same field at external points as if the charge were concentrated at its center; and the total negative charge on the remaining electrons in the atom is less by 1 electronic unit than the positive charge on the nucleus. If the field were everywhere of this simple Coulomb type, the energy levels for the valence electron would be exactly the same as those for hydrogen. Now the wave functions for large n and not too small l will have their large values chiefly in regions outside the atom, as is illustrated by the graphs for hydrogen in Fig. 89 in Sec. 124. Hence, the alkali levels for large n and not too small l should approximate closely to the hydrogen levels for the same n .

In Fig. 97 are shown the known energy levels below $2,000 \text{ cm.}^{-1}$ for all of the alkali metals, represented by dots. The hydrogen levels in the same spectral region (except the lowest) are shown by the horizontal lines labeled with the appropriate value of n . We note that the D and F levels of the alkalis actually agree closely in all cases with the corresponding hydrogen levels.

Such an approximate agreement with the hydrogen levels serves to fix the value of n unambiguously for the D and F levels of the alkalis. The quantum number n does not have any direct spectroscopic significance, hence the particular set of values that is used can only be chosen for some theoretical reason. The P and S levels, on the other hand, depart widely from the corresponding hydrogen levels. The lowest values of n for S and P terms are taken to be 2 for lithium, 3 for sodium, and so on, partly on the basis of the way in which the periodic table is built up, as was described previously in this chapter,

and partly on the basis of various theoretical calculations. When an approximate theory is developed with the electrons moving independently in a central field, the quantum number n is introduced in such a way as to number the 1-electron states in the order of increasing energy. With the values of n assigned as in the diagram, levels having the same n and L have the property that they sink as the atomic number increases. Theoretical estimates indicate that they should do this if n has the theoretical significance just stated. The 1-electron functions of the same n and l penetrate to about the same relative extent into the cores of all atoms, and, as the strength of the nuclear

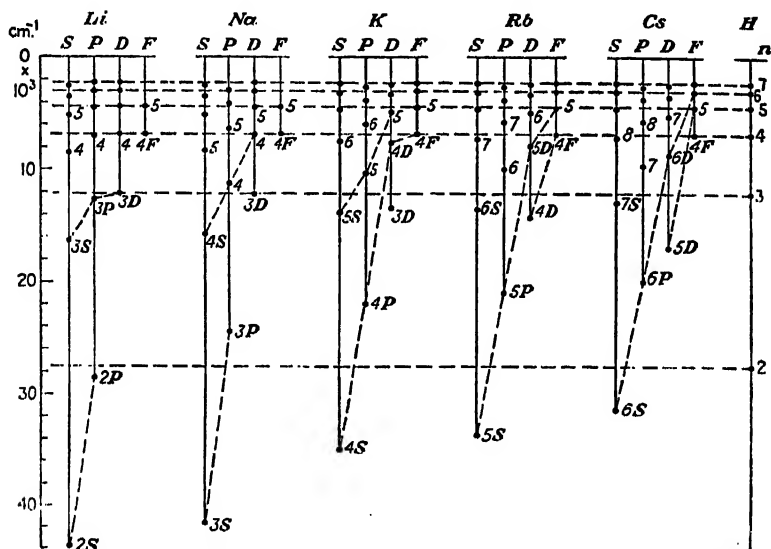


FIG. 97.—The known energy levels below 2,000 cm^{-1} of the five alkali metals and of hydrogen (except that for $n = 1$). All doublet structure is ignored.

field in the interior of the atom is increased with increasing Z , calculation shows that the corresponding 1-electron energy is lowered. The observed energies of the lowest S and P levels of the alkalis, on the other hand, *rise* with increasing Z . This fact is brought into consistency with the rule just stated if we suppose that n increases by unity from one alkali to the next.

Since the lowest S level is always below the lowest P , the normal state for all of the alkali metals is an S state.

The relation with the hydrogen levels just described suggests a mode of writing the alkali energies that is of interest. Expressed in wave-number units, the energy of the n th level for a hydrogen atom is

$$W_n^{(a)} = -\frac{R}{n^2},$$

where R is the Rydberg constant [cf. Eqs. (136), (140c) in Secs. 103 and 104]. For any atomic level with energy W , an *effective principal quantum number* n_e can be defined by the similar equation

$$W = -\frac{R}{n_e^2}. \quad (194)$$

The difference $n - n_e$ for corresponding levels is sometimes called the *quantum defect*. In the alkalis, it can be regarded as a measure of the effect of the departure of the central field from a Coulomb field. The quantum defect is often denoted by μ , so that

$$n_e = n - \mu;$$

then

$$W = -\frac{R}{(n - \mu)^2}. \quad (195)$$

A similar formula is obtained from Rydberg's expression for the spectral terms, number (129) in Sec. 97:

$$W = -\frac{R}{(m + \mu)^2}.$$

The values of μ are not the same in the two formulas, nor did Rydberg's choice of m for a given term agree with the modern value of n ; but, after choosing the lowest values of the integers n and of m in any convenient manner, the two values of μ can obviously be chosen so as to bring the two formulas into agreement. Since Rydberg's formula fits the observed terms fairly well, when his μ is treated as a constant for given L , it follows that the quantum defect must be almost a constant for the sequence of terms belonging to a given value of L .

The values of the quantum defect for some of the levels of sodium (disregarding the fine-structure splitting) are shown in Table IV.¹

TABLE IV.—SOME VALUES OF THE QUANTUM DEFECT FOR NEUTRAL SODIUM

Term	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
<i>S</i>	1.373	1.357	1.352	1.349	1.348	1.351
<i>P</i>	.883	.867	.862	.859	.858	.857
<i>D</i>	.010	.011	.013	.011	.009	.013
<i>F</i>000	— .001	— .008	— .012	— .015

¹ They are calculated as $n - n_e$ from values of n_e or n_{eff} given in H. E. White, "Introduction to Atomic Spectra," 1934.

Their approximate independence of n is evident. The corresponding value of the effective quantum number n_e may be obtained in each case by subtracting the number shown here from n ; thus for a $4P$ term, $n_e = 4 - 0.867 = 3.133$. A small value of the quantum defect means close approximation of the energy to the corresponding level for hydrogen.

Formulas of the Rydberg type, as in Eq. (195), represent only a first approximation to the actual term values. For greater accuracy more complicated functions of n are often employed in the denominator.¹ Such formulas are indispensable in determining series limits, and a knowledge of these limits is essential in order to fix the zero level of the energy, which is the limit of that series of lines

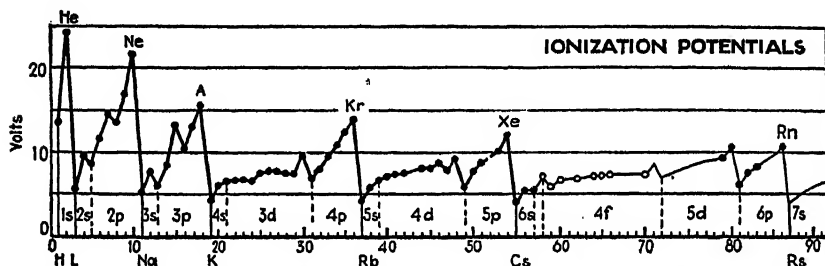


FIG. 98.—Ionization potentials as a function of atomic number. (Reprinted from H. E. White, "Introduction to Atomic Spectra.")

which arises from transitions into the normal state. Once the energies or term values have been determined for a single set of terms, the others can be calculated from the observed frequencies of lines due to transitions between the various terms.

The term value, or negative of the energy, for the *lowest state* represents the *ionization energy* of the neutral atom. In Fig. 98 are plotted values of the ionization energy calculated in this way and expressed in electron-volts for most of the elements. The electronic subshell that contains the most easily removed electron is shown below the curve. A systematic relation is evident between the quantum numbers of this electron and the ionization potentials.

144. The Addition of Orbital and Spin Momenta.—Besides accounting for the number and location of the terms, an atomic theory, to be successful, must lead also to the correct fine structure for the spectral lines. In order to account for the fine structure it is necessary to reintroduce the spin-orbit terms into the wave equation. Their effect, as it turns out, is to diminish the degeneracy of the atomic

¹ Cf. WHITE, "Introduction to Atomic Spectra," 1934; FOWLER, A., "Report on Series in Line Spectra," Fleetway Press, London, 1922.

quantum states and to split most of the LS terms into two or more levels characterized by different values of J , the quantum number for the total angular momentum. It will be convenient to discuss this splitting first for the general case, thereby obtaining results applicable to an atom containing any number of active electrons. As usual, we shall merely describe the theoretical method very briefly and write down some of the results.

Following the method of perturbation theory, the theorist starts with a wave equation from which the spin-orbit terms are omitted. Using this wave equation, as stated above (Sec. 141), we can find a set of quantum states for the atom characterized by four quantum numbers $L \Lambda S \Sigma$. Consider such a set for a given L and a given S . The possible values of Λ are $2L + 1$ in number (from L to $-L$), and there are $2S + 1$ different possible values of Σ ; thus, we have $(2L + 1)(2S + 1)$ states with the same L and S . These states all have the same energy, also, in this stage of the argument; let us denote this value of the energy by W_0 .

The spin terms are then restored to the wave equation. These terms do not alter the number of the quantum states, but they do change the energies and the wave functions. If their effect is small, there will exist a set of states for the given L and S having energies near to W_0 . The wave functions for these states do not resemble the functions for the original $L \Lambda S \Sigma$ states, however; they resemble certain linear combinations of the latter functions.

An excellent analogy occurs in classical vibrations. A pendulum bob hung on a symmetrical wire can swing with a certain definite frequency ν_0 in any plane. It can swing east and west, for example, or north and south; and vibration in any other plane can be regarded as a combination of these two motions, each with a suitable amplitude and phase. Any other pair of perpendicular planes might be chosen instead for the component vibrations. Such a case may be called "degenerate." But now squeeze the wire with pliers so as to flatten it in a NE - SW direction. Thereafter, steady oscillations in a fixed plane are possible in only two directions. The bob can swing either in a NE - SW plane with a frequency ν_1 or in a NW - SE plane with a frequency ν_2 , ν_1 and ν_2 being slightly different from each other and from ν_0 . If the bob is started swinging in any other direction, a combination of these two vibrations occurs and the bob executes a Lissajous figure. Thus flattening the wire has removed the degeneracy.

In a similar way, there is an ambiguity about the functions for the quantum states when the spin-orbit terms are missing from the wave equation; functions characterized by quantum numbers $L S \Lambda \Sigma$ repre-

sent only one possible choice for these functions. An alternative is to construct the functions so that they are characterized by quantum numbers

$$L S J M.$$

If the latter is done, each of the functions represents the atom as having orbital angular momentum $\sqrt{L(L+1)} h/2\pi$, also spin angular momentum $\sqrt{S(S+1)} h/2\pi$, and, finally, total angular momentum $\sqrt{J(J+1)} h/2\pi$ and also a component of total momentum about any chosen axis equal to $Mh/2\pi$. But the components about this axis of the orbital and spin momenta, taken separately, are no longer fixed. In the absence of the spin-orbit terms all of these states, like the $L S \Lambda \Sigma$ states, have the same energy.

When the *spin-orbit terms* are then restored to the wave equation, they do not destroy the physical significance of the quantum numbers J and M ; and, if the spin-orbit effect is small, the significance of L and S still remains, in large measure. The energies of levels with the same values of L and S but with different values of J now become somewhat different, however. Thus the spin-orbit effect partially removes the degeneracy. For each value of J , there are still $2J + 1$ states with various values of M . We shall find later that these latter states, also, can be separated in energy by the application of a magnetic field.

Two questions then arise: (1) How many and what values of J occur with a given pair of values of L and S , and (2) how far are the J levels separated from each other? The answers to these questions will be discussed in the next two sections.

145. Number of Levels in an LS Multiplet.—The group of energy levels corresponding to different values of J but to the same values of L and S is called an " LS " (or "*Russell-Saunders*") *multiplet*. For the number of the levels composing such a multiplet, wave-mechanical theory furnishes a simple rule. The maximum value of J that occurs is $J = L + S$. Then, if neither L nor S is zero, $J = L + S - 1$ occurs; also $J = L + S - 2$ if this number is not negative; and so on down to $J = |L - S|$ as the smallest value. Thus, a level occurs with every one of the set of integrally spaced values of J such that

$$|L - S| \leq J \leq L + S. \quad (196)$$

Thus, if $L \geq S$, J takes on all of the $2S + 1$ integrally spaced values from $L + S$ down to $L - S$, inclusive; if $L < S$, it takes on the $2L + 1$ integrally spaced values from $L + S$ down to $S - L$. That the greatest and least values of J should be, respectively, $L + S$ and $|L - S|$ follows at once from classical ideas concerning vector addi-

tion; but that the possible values of J should be limited to a discrete set constitutes a characteristic quantum modification.

Since L is always an integer or zero, the values of J , like those of S , will obviously be integral if the number of electrons in the atom is even, half-integral if this number is odd. For example, if $L = 0$ and $S = 0$, then $J = 0$; if $L = 0$, $J = S$; if $S = 0$, $J = L$; if $L = 1$ and $S = \frac{1}{2}$, $J = \frac{3}{2}$ or $\frac{1}{2}$ (i.e., $L \pm S$); if $L = 1$ and $S = 1$, $J = 2$, 1, or 0.

These relationships can be visualized by means of a *vector diagram* representing an L -vector and an S -vector added vectorially into a resultant J -vector. In the sense explained in connection with the J -vector in Sec. 140, the L - and S -vectors can be thought of as representing the total orbital and total spin momenta respectively. The

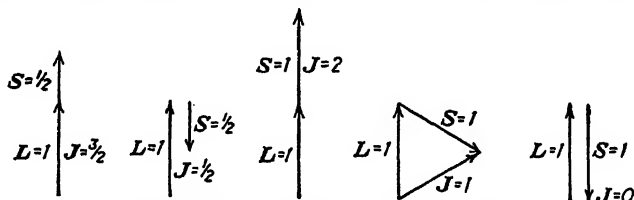


FIG. 99.—Diagrams illustrating the relation between L , S , and J .

value $J = L + S$ corresponds to addition of these two momenta in the same direction, the value $J = |L - S|$, to their addition in opposite directions; every other value of J corresponds to their addition at one of a discrete set of intermediate angles. [If vectors of lengths $\sqrt{L(L+1)}$, $\sqrt{S(S+1)}$, and $\sqrt{J(J+1)}$ are drawn, in order to represent the vectors more completely in accordance with the explanation in Sec. 140, the angle between L and S is less than 180° even for $J = L + S$ and $J = |L - S|$.] A few such vector diagrams are shown in Fig. 99.

In discussing LS coupling, it is often convenient to restrict the word "term" so as to refer to a group of levels with various J values forming a multiplet. From the rule just stated for determining the possible values of J , it is evident that, if $L \geq S$, the number of levels in the multiplet is $2S + 1$. This number is called the *multiplicity* of the LS term and is written as a superscript at the upper left-hand corner of the letter denoting the term. Thus 2D denotes a doublet D term, with $L = 2$, $S = \frac{1}{2}$ (i.e., $2 = 2 \times \frac{1}{2} + 1$), and with two possible J values: $L + S = \frac{5}{2}$, $L - S = \frac{3}{2}$. To denote an individual level, the J value is added as a subscript. Thus a 2D term contains the levels ${}^2D_{\frac{5}{2}}$, ${}^2D_{\frac{3}{2}}$. If $L = 2$, but $S = 1$, we have a *triplet* D term, with component levels 3D_3 , 3D_2 , 3D_1 .

For convenience, even when $L < S$, the same multiplicity, $2S + 1$, is indicated in the notation, although the number of levels now equals $2L + 1$ and therefore is less than $2S + 1$. The reason for this practice is that, because of the selection rule for S , $\Delta S = 0$ (Sec. 141), the value of $2S + 1$ tends to remain the same in a radiative transition. The terms can thus be classified into more or less noncombining groups according to the values of $2S + 1$, and the superscript serves to indicate at once to which class a particular term belongs.

Because of these relations, the *observed* number of levels composing a given multiplet obviously furnishes important information in regard to the values of L or of S that should be assigned to the term. If the observed number of levels is τ , this number must equal either $2L + 1$ or (more commonly) $2S + 1$, hence, either $L = (\tau - 1)/2$ or, more likely, $S = (\tau - 1)/2$. A further test may then be made by noting whether or not the value of L or S so inferred is consistent with the selection rules as applied to transitions between the given multiplet of levels and other multiplets for which L or S may be assumed to be known.

All allowed transitions between two multiplets of levels, taken together, give rise to a group of lines which may be called a *spectral multiplet* of lines. The student should distinguish carefully between multiplets of levels and multiplets of lines. Often one can tell only from the context whether the word "triplet," for example, refers to three energy levels or to three spectral lines.

In Table V are shown the levels composing some of the types of LS terms that are commonly met with. Only singlet, doublet, triplet, and

TABLE V.— J LEVELS COMPOSING SOME LS TERMS

$L =$	0	1	2	3
$S = 0$	1S_0	1P_1	1D_2	1F_3
$\frac{1}{2}$	$^2S_{\frac{1}{2}}$	$^2P_{\frac{1}{2}}, ^2P_{\frac{3}{2}}$	$^2D_{\frac{3}{2}}, ^2D_{\frac{5}{2}}$	$^2F_{\frac{5}{2}}, ^2F_{\frac{7}{2}}$
1	3S_1	$^3P_0, ^3P_1, ^3P_2$	$^3D_1, ^3D_2, ^3D_3$	$^3F_2, ^3F_3, ^3F_4$
$\frac{3}{2}$	$^4S_{\frac{3}{2}}$	$^4P_{\frac{1}{2}}, ^4P_{\frac{3}{2}}, ^4P_{\frac{5}{2}}$	$^4D_{\frac{1}{2}}, ^4D_{\frac{3}{2}}, ^4D_{\frac{5}{2}}, ^4D_{\frac{7}{2}}$	$^4F_{\frac{3}{2}}, ^4F_{\frac{5}{2}}, ^4F_{\frac{7}{2}}, ^4F_{\frac{9}{2}}$

quartet terms are shown, but quintet, sextet, septet, and octet terms also occur.¹

As stated above, when the spin-orbit effect is rather large, the significance of L and S becomes weakened, and the selection rules tend to fail; in heavy atoms, indeed, it may be impossible to assign values of L and S at all, so that another grouping of the J levels must be made

¹ Cf. WHITE, *op. cit.*, Chap. XIV, 1934.

(e.g., that represented by *jj*-coupling as described in Sec. 153 below). Thus the quantum number J possesses a wider range of significance than the numbers L and S .

146. Spin-orbit Interaction and Multiplet Spacing.—To obtain theoretical predictions as to the *spacing* or *energy differences* between the levels of a multiplet, it is necessary to consider the effect of the spin-orbit terms in the wave equation. A more detailed discussion of these terms is appropriate at this point.

The spin-orbit terms put in their appearance when a *nonrelativistic* equation is worked out that is approximately equivalent to the Dirac *relativistic* equation, which is the nearest approach known to a correct wave equation (Sec. 126). The Dirac equation does not rest on a concrete picture of the electron. It seems as if theoretical physics were coming to be based upon certain mathematical assumptions rather than upon concrete pictures of reality. There is, however, a certain similarity between the wave-mechanical mathematics and the mathematical treatment of certain classical structures, and this analogy may help by illustrating, and perhaps making plausible, the more abstract treatment by means of wave mechanics.

Suppose the electron really were, as Lorentz thought it might be, a little hollow sphere of electricity, held together in some unknown manner but otherwise obeying the laws of classical theory, relativity included. Suppose the sphere were set into rotation. Then the moving charges on its surface would give rise to circular currents, and the sphere would be surrounded by a magnetic field resembling that due to a short bar magnet; it would, therefore, have a certain magnetic moment. Suppose such a rotating sphere were put in motion across an electric field. Then ordinary electromagnetic theory would not lead us to expect any special forces to act on the sphere because of its magnetic moment. But suppose we view the situation from a moving frame of reference, in which the sphere is momentarily at rest, and apply the theory of relativity. In terms of this moving frame, the electric field would be accompanied by a magnetic field perpendicular both to the electric field and to the direction of motion, as is shown by the equations of transformation in Sec. 70.¹ This magnetic field would cause the magnetic axis of the sphere to tend to line up with the field.

¹ In those equations, we put $H_x = H_y = H_z = 0$, so that in S there is only an electric field. Then $H'_x = 0$, showing that the magnetic field in S' is perpendicular to the x -axis, which is the direction of the motion. Each of the components E_y and E_z then gives rise to a component of H' equal to itself multiplied by u/c , but in a perpendicular direction.

Returning to the original frame, therefore, we may conclude that any rotating charged sphere moving transversely across an electric field tends to turn so as to make its axis of rotation perpendicular both to its direction of motion and to the direction of the field. If the motion occurs in the direction of the field, or if there is no field at all, the axis of rotation retains its original direction. As viewed in this frame, the immediate cause of the change in the direction of the axis is to be found in a relativistic change in the forces that hold the charged sphere together against electrostatic repulsion.

We do not need to investigate these forces further, but the source of the energy associated with them is of interest and is easily discovered. When the sphere is aligned in the manner just described (Fig. 100), one side of it is moving faster than the other side because of the combined rotational and translational motion. Now in a frame in which the sphere is at rest, the rotational velocity of all parts of its surface must be the same, by symmetry; but then an application of the Lorentz transformation for velocity, Eqs. (63*a,b,c*) in Sec. 65, shows that, in a frame in which the sphere has translational motion, the surface will rotate more slowly on the side on which the resultant velocity is increased by the translational motion than on the opposite side. This difference in velocities must result in a piling up of charge on the side of higher resultant velocity at the expense of charge on the opposite side.¹ If the sphere is moving across an electric field, the distortion shifts the charge into a region of slightly different potential and so alters the electrostatic energy. A simple consideration of signs shows that, in the case of orbital motion of the sphere under the influence of an attracting center, the energy is least when the direction of spin rotation is opposite to the direction of orbital revolution and is a maximum when these two directions of rotation are the same, regardless of the sign of the charge on the sphere.

The magnitude of the change in energy for a Lorentz electron can readily be calculated, but we shall only cite the result.² In dealing with an electron in a central field, it is convenient to introduce the angular momentum G_L of the electron about the center. Let G_S be its angular momentum of spin, m its mass, c the speed of light, and

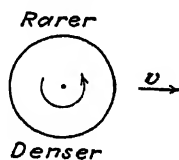


FIG. 100.—A classical (Lorentz) spinning electron viewed along its axis of rotation.

¹ The distortion of the distributed charge can be regarded as equivalent to the creation of an electric moment relative to the center of charge. Thus what is a pure magnetic moment in one frame of reference becomes in part an electric moment in another.

² FRENKEL, J., "Lehrbuch der Elektrodynamik," vol. II, Kap. 7, §8, 9, 1926.

$V(r)$ its potential energy when at a distance r from the center of the field. Then the change in energy due to interaction of spin and field turns out to be

$$\frac{1}{2m^2c^2} \mathbf{G}_L \cdot \mathbf{G}_S \frac{1}{r} \frac{dV}{dr}.$$

Here $\mathbf{G}_L \cdot \mathbf{G}_S$ denotes the scalar product of the vectors \mathbf{G}_L and \mathbf{G}_S or

$$\mathbf{G}_L \cdot \mathbf{G}_S = G_L G_S \cos (G_L, G_S) \quad (197)$$

where G_L and G_S denote the magnitudes of these vectors and (G_L, G_S) stands for the angle between their directions.

The expression thus obtained from the classical analogy happens to be correct in wave mechanics also, although the physical picture of the rotating sphere finds no place in the new theory. The spin-orbit term in the wave equation for 1 electron, in the approximate non-relativistic theory, is

$$\frac{1}{2m^2c^2} \frac{1}{r} \frac{dV}{dr} \mathbf{G}_L \cdot \mathbf{G}_S \psi, \quad (198)$$

in which \mathbf{G}_L and \mathbf{G}_S stand for vector "operators." (An operator is a quantity that does certain specified things to any function standing after it.) The first-order correction to the energy caused by the spin-orbit effect is then, by the usual formula of perturbation theory,

$$\Delta W = \frac{1}{2m^2c^2} \int \int \int \psi^* \frac{1}{r} \frac{dV}{dr} \mathbf{G}_L \cdot \mathbf{G}_S \psi dx dy dz, \quad (199)$$

the integral being extended over all space and ψ^* denoting the complex conjugate of ψ . If the field is a Coulomb field, $V = -Ze^2/r$ (Z = atomic number, e = electronic charge) and (199) becomes

$$\Delta W = \frac{Ze^2}{2m^2c^2} \int \int \int \frac{1}{r^3} \psi^* \mathbf{G}_L \cdot \mathbf{G}_S \psi dx dy dz. \quad (199a)$$

The expression given in Eq. (199) for ΔW can be evaluated by the theorist if ψ and V are known. If LS coupling holds, furthermore, the most interesting feature of ΔW can be inferred without evaluating the integral. The necessary argument can be illustrated by means of the vector diagram. Let us draw lines representing as closely as possible, not the quantum numbers L and S , but the resultant orbital momenta themselves, \mathbf{G}_L and \mathbf{G}_S , added vectorically to form the resultant angular momentum \mathbf{G}_J (Fig. 101). Let us draw the lines of

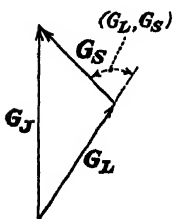


FIG. 101.

such length as to represent the true magnitudes of the three momenta, which are

$$G_L = \sqrt{L(L+1)} \frac{h}{2\pi}, \quad G_S = \sqrt{S(S+1)} \frac{h}{2\pi},$$

$$G_J = \sqrt{J(J+1)} \frac{h}{2\pi}.$$

Then by ordinary geometry

$$G_J^2 = G_L^2 + G_S^2 + 2G_L G_S \cos(G_L, G_S).$$

Combining this relation with (197), we obtain

$$\mathbf{G}_L \cdot \mathbf{G}_S = \frac{1}{2}(G_J^2 - G_L^2 - G_S^2). \quad (200)$$

The same relation is easily shown to hold in wave mechanics. Hence, after inserting the values of G_L^2 , G_S^2 , G_J^2 ,

$$\mathbf{G}_L \cdot \mathbf{G}_S = \frac{1}{2}[J(J+1) - L(L+1) - S(S+1)] \frac{h^2}{4\pi^2}.$$

Thus when ψ represents a quantum state characterized by quantum numbers L , S and J , the operator $\mathbf{G}_L \cdot \mathbf{G}_S$ acting on ψ is equivalent to a simple multiplication of it by the number given on the right in the last equation. This number can then be taken out from under the integral sign in Eq. (199), and that equation can be written

$$\Delta W = \frac{1}{2}B[J(J+1) - L(L+1) - S(S+1)], \quad (201)$$

$$B = \frac{h^2}{8\pi^2 m^2 c^2} \int \int \int \frac{1}{r} \frac{dV}{dr} |\psi|^2 dx dy dz. \quad (201a)$$

Here the integral in B represents a sort of weighted average of the quantity $(1/r) (dV/dr)$, which can be calculated if ψ is known. For ψ , we may use the zero-order wave function for an $LSJM$ state. It can be shown that, for given L and S , B is independent of J and M .

The spacing of the J levels of an LS multiplet will thus be determined by that term in Eq. (201) which contains $J(J+1)$, this being the only term that varies with J . The difference between the energies of a level for J and one for $J+1$ will be the same as the difference in the corresponding values of ΔW , or

$$W_{J+1} - W_J = \frac{1}{2}B(J+1)(J+2) - \frac{1}{2}BJ(J+1)$$

$$= B(J+1). \quad (202)$$

This equation includes Landé's interval rule: *The energy differences between two successive J levels are proportional, in a given term, to the larger of the two values of J .* This rule is of great help in determining the value of J that is to be assigned to the various levels.

It may be remarked that, if we employ the whole expression for ΔW as given by Eq. (201), we find that ΔW is positive for some values

of J and negative for others. As a matter of fact, the weighted average of ΔW , or $\Delta \bar{W}$ for a weighted-average level, computed in a certain way,¹ vanishes.

It is to such a weighted-average level that the Rydberg formula really refers when it is written without regard to the fine structure of the levels. As an alternative, of course, a separate Rydberg formula can be written for each component of the multiplets of a series, as is sometimes done.

147. Fine Structure in Alkali-type Spectra.—The theory developed in the last section is immediately applicable to energy levels of the type described in Sec. 142. For an atom with 1 valence electron outside of one or more closed subshells, the quantum number S has the value $S = s = \frac{1}{2}$, as for 1 electron (Sec. 126), for the same reason

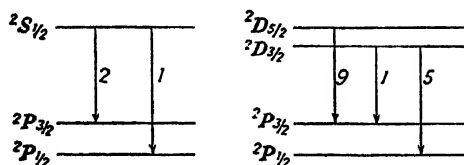


FIG. 102.—Transitions between doublet levels.

which led us in Sec. 142 to infer that $L = l$. Hence, the LS terms for such an atom are mostly *doublets*, the two values of J for given L being $J = L + \frac{1}{2}$ and $J = L - \frac{1}{2}$. Only the S terms ($L = 0$) are single, with $J = S = \frac{1}{2}$. Almost always the level with the larger J is found to have the higher energy, in agreement with indications from wave mechanics.

Transitions between two given terms will then give rise to several spectral lines, forming a spectral multiplet. The number of the lines is limited by the selection rule, $\Delta J = 0$ or ± 1 (Sec. 140). Jumps are possible between the single level of any S term, $2S_{1/2}$, and either of the two levels in a P term, $2P_{1/2}$ and $2P_{3/2}$, for which $\Delta J = 0$ or 1, respectively. Between a P and a D , with levels $2D_{3/2}$, $2D_{5/2}$, three jumps are possible, *viz.*,

$$2D_{3/2} - 2P_{1/2}, \quad 2D_{3/2} - 2P_{3/2}, \quad 2D_{5/2} - 2P_{3/2};$$

the jump $2D_{5/2} - 2P_{1/2}$ is forbidden, since for it $\Delta J = 2$. A triplet of lines results in a similar way from the allowed transitions between any other pair of terms. Thus *the chief spectrum of the alkali metals*

¹ If W_J denotes the energy of a J level, the energy of the weighted-average level is $\bar{W} = \sum_J (2J + 1) W_J / \sum_J (2J + 1)$. (There are $2J + 1$ quantum states in each J level.)

should consist of spectral doublets and triplets. The origin of the lines just mentioned is illustrated in Fig. 102. Numbers are also added in the figure showing the relative intensities that the lines of each

TABLE VI.—SOME LINES IN THE SPECTRUM OF NEUTRAL SODIUM

Initial term, cm. ⁻¹	$\bar{\nu}$, cm. ⁻¹ *	λ , angstroms	Doublet $\Delta\bar{\nu}$	
Principal series, $n^2P \rightarrow 3^2S$. $3S = 41,449.0$ cm. ⁻¹				
$3P_{1/2}$	24,492.8	16,956.19	5,895.92	17.19
$3P_{3/2}$	24,475.6	16,973.38	5,889.95	
$4P_{1/2}$	11,182.1	30,266.91	3,302.99	6.15
$4P_{3/2}$	11,175.9	30,273.06	3,303.32	
$5P_{1/2}$	6,406.4	35,042.64	2,852.83	2.46
$5P_{3/2}$	6,408.8	35,040.18	2,853.03	
$6P_{1/2}$	4,151.4	37,297.64	2,680.34	1.40
$6P_{3/2}$	4,152.8	37,296.24	2,680.44	
Sharp series, $n^2S \rightarrow 3^2P$.				
$4S$	15,709.4	8,766.5	11,404.2	16.6
		8,783.1	11,382.4	
$5S$	8,248.3	16,227.28	6,160.76	17.22
		16,244.50	6,154.23	
$6S$	5,077.3	19,398.38	5,153.64	17.13
		19,415.51	5,149.09	
$7S$	3,437.2	21,038.40	4,751.80	17.14
		21,055.54	4,748.02	
Diffuse series, * $n^2D \rightarrow 3^2P$.				
$3D$	12,276.12	12,199.50	8,194.81	17.20
		12,216.70	8,183.27	
$4D$	6,900.29	17,575.33	5,688.22	17.19
		17,592.52	5,682.66	
$5D$	4,412.3	20,063.30	4,982.84	17.16
		20,080.46	4,978.58	
$6D$	3,061.9	21,413.72	4,668.60	17.18
		21,430.90	4,664.86	

* Only the mean value of the $D_{5/2}$ and $D_{3/2}$ levels is shown and only the two strong lines,

$$nD_{5/2} \rightarrow 3P_{3/2}, \quad nD_{3/2} \rightarrow 3P_{1/2}.$$

multiplet should have according to wave mechanics, under certain simple conditions.

For comparison with the theory, the wave numbers and separations of the first four lines of each of the chief series in the spectrum of neutral sodium are listed in Table VI.¹

The theoretical conclusions as to wave length are seen to be confirmed by the observations. The *sharp* series, $nS \rightarrow 3P$, consists of spectral doublets, and, as shown in column 5 of Table VI, the frequency difference is constant within the experimental errors, representing the difference between the $3^2P_{3/2}$ and $3^2P_{1/2}$ levels. This fact helps to confirm the identification of the S levels as made above.

The *principal* series of lines, $nP \rightarrow 3S$, also consists of doublets but with frequency differences that represent the spacings between the two levels in various P terms and hence decrease rapidly with increasing n . This decrease is in accord with theoretical expectations. For an electron in a Coulomb field, due to a nucleus of charge Ze , the separation of the two levels in a term with quantum numbers n and l , due to the spin-orbit effect, is readily calculated to be [cf. Eqs. (208a,b) below]

$$\Delta\bar{\nu} = \frac{\alpha^2 R Z^4}{n^2 l(l+1)} = 5.87 \frac{Z^4}{n^2 l(l+1)} \text{ cm.}^{-1} \quad (203)$$

where α is the "fine-structure constant," or $\alpha = 2\pi c^2/ch = 1/136.7$, and R is the Rydberg constant (in cm.^{-1}). The calculation can be made by evaluating the integral in the expression given for B in Eq. (201a), with $V = -Ze^2/r$, and then substituting the value so found for B and also $J = l - \frac{1}{2}$ in Eq. (202). Dividing the resulting value of $W_{J+1} - W_J$, or $W_{l+\frac{1}{2}} - W_{l-\frac{1}{2}}$, by hc , we have $\Delta\bar{\nu}$. Thus in a hydrogenlike atom the spin-orbit separation decreases as $1/n^3$. The observed decrease in the spectrum of sodium is even more rapid.

The lines of the *diffuse* series, $nD \rightarrow 3P$, should be *triplets*, according to the theory. One line should be very weak, however. The two brightest lines result from the jumps

$$n^2D_{5/2} \rightarrow 3^2P_{3/2}, \quad n^2D_{3/2} \rightarrow 3^2P_{1/2}.$$

If Eq. (203) holds roughly for sodium, then, because of the factor $l(l+1)$ in the denominator, the separation $nD_{5/2} - nD_{3/2}$ should stand to the separation $nP_{3/2} - nP_{1/2}$ in the ratio $1/(2 \times 3):1/(1 \times 2)$ or as only 1:3; and, as n increases, the D separation should rapidly diminish further. Thus with ordinary resolving power the lines of the

¹ Except for $\lambda 11,404$, or $\lambda 11,382$, the wave lengths of the lines were taken from the "Massachusetts Institute of Technology Wavelength Tables" and converted into wave numbers with the help of Kayser's "Tabelle der Schwingungszahlen"; the term values were then calculated from the wave numbers, the value of the $3S$ term (41,449.0) being taken from Bacher and Goudsmit's "Atomic Energy States."

diffuse series can easily be observed with a frequency difference that is nearly constant and equal to the difference $3P_{3/2} - 3P_{1/2}$. Doublets of *exactly* this separation are formed of course by the faint line $D_{3/2} \rightarrow 3P_{3/2}$ and the brighter line $D_{3/2} \rightarrow 3P_{1/2}$. The faint line is not usually seen; when it was discovered, it was called a "satellite" of the apparent doublet formed by the two brighter lines.

Thus, the theory accounts very well for the principal features of the spectrum emitted by neutral sodium atoms. It is equally successful with the other alkali metals, the spectra of which are qualitatively very similar to that of sodium.

The separation of the doublet levels, however, increases rapidly with increasing Z . In Table VII are shown the wave lengths and the wave numbers of the D lines or their analogs (*i.e.*, the first lines of the principal series) for all of the alkali metals, and also the doublet differences for their lines.

TABLE VII.—FIRST LINES OF THE PRINCIPAL SERIES FOR THE ALKALI METALS, AND ONE LINE OF HYDROGEN
 λ is in angstroms, $\tilde{\nu}$ in cm^{-1}

	H	Li	Na	K	Rb	Cs
Z	1	3	11	19	37	55
n	3 \rightarrow 2	2	3	4	5	6
$\lambda_{1/2}$			5,895.9	7,699.0	7,947.6	8,943.5
$\lambda_{3/2}$	6,562.8	6,707.8	5,890.0	7,664.9	7,800.2	8,521.1
$\tilde{\nu}_{1/2}$			16,956	12,985	12,579	11,178
$\tilde{\nu}_{3/2}$	15,233	14,904	16,973	13,043	12,817	11,732
$\Delta\tilde{\nu}$	0.367	0.34	17	58	238	554

The theoretical value of $\Delta\tilde{\nu}$ for the first Balmer line of hydrogen, obtained by setting $Z = 1$, $n = 2$, $l = 1$ in Eq. (203), is also given for comparison. The data in the table show that the D lines, which are 6 angstroms apart in the sodium spectrum, are 422 angstroms apart in the spectrum of cesium.

The enormous departure of $\Delta\tilde{\nu}$ from the hydrogen value for all alkali metals except lithium, in spite of the progressive increase in n , is in strong contrast with the more moderate departure of the energy levels themselves, as shown by Fig. 97. This may be regarded as resulting from great sensitiveness of the spin-orbit effect to the character of the central field near the nucleus, which is clearly evident in the expression written down above, (198) in Sec. 146, for the spin-orbit term. Near the nucleus the field of the nucleus itself must predomi-

nate, so that approximately $V = -Ze^2/r$. Thus, the potential term in the wave equation would be $V\psi = -Ze^2\psi/r$, whereas in (198) we have

$$\frac{1}{r} \frac{dV}{dr} = \frac{Ze^2}{r^3},$$

which increases much more rapidly than does Ze^2/r as the nucleus is approached. The variation of the spin-orbit effect among the alkalis furnishes an excellent illustration of the tendency of this effect to be small in atoms of low atomic number but to increase to an enormous magnitude as the atomic number becomes large.

148. Multiplet Levels for One-electron Atoms.—In the last section, a theory of the fine structure of spectral lines due to spin was developed for atoms containing a valence electron outside of closed subshells. The same theoretical treatment should be applicable also

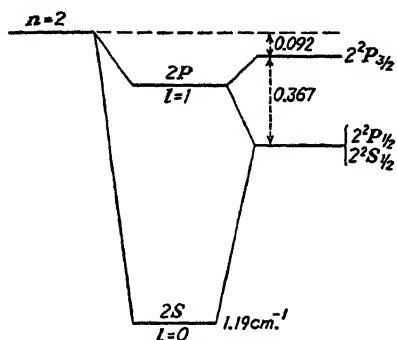


FIG. 103.—The fine structure for $n = 2$ for ordinary hydrogen: at left, nonrelativistic theory; center, as modified by the ordinary relativistic correction; right, actual arrangement of levels.

to atoms containing no subshells at all, *i.e.*, to atoms such as those of hydrogen which contain *only* a single electron. In this latter case, it is customary to use small letters, $l\ j\ m$ (or m_j) instead of $L\ J\ M$; then $j = \frac{1}{2}$ if $l = 0$, otherwise, $j = l + \frac{1}{2}$ or $l - \frac{1}{2}$.

When only 1 electron is present, however, a curious accident occurs. In the nonrelativistic theory, as we have seen in Sec. 124, all states for a given n have the same energy, regardless of the value of l . In the relativistic theory this is not quite

true. The wave equation contains certain other small terms of relativistic origin, besides those giving rise to the spin-orbit effect, and these other terms cause the energy to vary somewhat with l . In a 1-electron atom this variation happens to be of the same order of magnitude as the spin-orbit effect itself; in fact, the net result is that levels having different values of l but the same value of j fall on top of each other. The energy diagram for such atoms is thus very peculiar. Somewhat similar relativistic effects occur also in atoms containing 2 electrons (and in them the interaction between the spins of the 2 electrons is important); but in such cases the resulting displacement of the various L terms is much smaller than is their displacement due to the electrostatic repulsion between the electrons.

The situation in the 1-electron atoms is illustrated in Fig. 103, which refers to the hydrogen levels with $n = 2$. The single line at the left represents the energy as given by Bohr's theory, Eq. (136) in Sec. 103. The two solid lines in the center of the diagram show this level as split into two, a P and an S term, in consequence of the "ordinary" relativistic correction. The addition of the spin-orbit effect then splits the $2P$ term into the usual doublet with $j = \frac{1}{2}$ or $\frac{3}{2}$; but an additional "special" spin correction raises the $2S$ level so that it coincides with the $2P_{\frac{1}{2}}$ level. The net result, as shown at the right in the figure, is just *two separate levels*, one with $j = \frac{3}{2}$ and a composite level with $j = \frac{1}{2}$. In a similar way, the $3S_{\frac{1}{2}}$ level comes to coincide with the $3P_{\frac{1}{2}}$ level, and $3P_{\frac{3}{2}}$ with $3D_{\frac{3}{2}}$, whereas $3D_{\frac{5}{2}}$ stands alone; and so on. Thus for each value of n there are n different energy levels, with $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, n - \frac{1}{2}$; but all levels except that for the largest j , $j = n - \frac{1}{2}$, are double. The lowest level of all, with $n = 1$, is still strictly single.

Exact formulas for the energy levels can be obtained by solving the Dirac relativistic wave equation (Secs. 126, 146). The result for the energy of a level specified by quantum numbers n and j is¹

$$W_{nj} = mc^2 \left\{ \left[1 + \frac{\alpha^2 Z^2}{(n - j - \frac{1}{2} + \sqrt{(j + \frac{1}{2})^2 - \alpha^2 Z^2})^2} \right]^{-\frac{1}{2}} - 1 \right\}.$$

Here α is the "fine-structure" constant:

$$\alpha = \frac{2\pi e^2}{ch} = 0.007314 = \frac{1}{136.7}. \quad (204)$$

This happens to be exactly the formula obtained by Sommerfeld on the basis of the old quantum theory, assuming elliptical orbits but allowing for the variation of electronic mass with speed, except that he used a different set of quantum numbers. His theory has been superseded by the wave-mechanical treatment; a short description of it was given in Sec. 111.²

For most purposes, it is sufficient to expand the expression written above in powers of α^2 and to keep only the first two terms of the expansion. By patiently applying the binomial theorem to one radical or fraction after another, one finds eventually that

$$W_{nj} = -\frac{mc^2 \alpha^2 Z^2}{2n^2} + \frac{mc^2 \alpha^4 Z^4}{2n^3} \left(\frac{3}{4n} - \frac{1}{j + \frac{1}{2}} \right) \dots$$

¹ Cf. DARWIN, C. G., *Roy. Soc., Proc.*, vol. 118, p. 654 (1928).

² Cf. SOMMERFELD, A., "Atomic Structure and Spectral Lines," Methuen & Co., Ltd., London, 1929.

ergs. Let us divide this by hc to convert it into wave-number units and then introduce the Rydberg constant or [cf. Eq. (146b) in Sec. 106]

$$R = \frac{2\pi^2 me^4}{ch^3} = \frac{mc\alpha^2}{2h} \quad (205)$$

by (204). Then, in wave-number units, as far as terms in α^2 ,

$$W_{nj} = -\frac{RZ^2}{n^2} + \frac{\alpha^2 RZ^4}{n^3} \left(\frac{3}{4n} - \frac{1}{j + \frac{1}{2}} \right). \quad (206)$$

The first term in this expression for W_{nj} is the ordinary nonrelativistic value of the energy [Eq. (186) in Sec. 124]. The second term represents the combined relativity-spin correction; it cannot exceed in magnitude a fraction α^2 or 5.35×10^{-5} of the first. The part of the second term that varies with j can be written

$$\Delta_i W = -\frac{\alpha^2 RZ^4}{n^3(j + \frac{1}{2})} = -5.87 \frac{Z^4}{n^3(j + \frac{1}{2})} \text{ cm}^{-1}. \quad (206a)$$

We note that this term in the energy, which determines the fine structure, decreases rapidly with n .

The *approximate* formula (206) is easily obtained by making in succession the two corrections mentioned above, with the help of perturbation theory. It may be of interest to write down the results of the successive steps. The *ordinary relativistic* correction to the energy is found to be

$$\Delta_1 W = \frac{3}{4} \frac{\alpha^2 RZ^4}{n^4} - \frac{\alpha^2 RZ^4}{n^3(l + \frac{1}{2})}. \quad (207)$$

This happens to be the same as Sommerfeld's correction for variable mass. It causes the splitting of each level for given n into n sublevels with $l = 0, 1, \dots, n-1$. The *spin-orbit* correction is then found (*e.g.*, by evaluating the integral in (201a) above) to be, for $l > 0$:

$$j = l + \frac{1}{2}: \quad \Delta_s W = \frac{1}{l+1} \frac{\alpha^2 RZ^4}{n^3(2l+1)}; \quad (208a)$$

$$j = l - \frac{1}{2}: \quad \Delta_s W = -\frac{1}{l} \frac{\alpha^2 RZ^4}{n^3(2l+1)}. \quad (208b)$$

For $l = 0$, the spin-orbit effect vanishes. The *special spin* correction is likewise found to vanish for $l > 0$, but for $l = 0$ it is

$$\frac{\alpha^2 RZ^4}{n^3}$$

This latter expression happens to be exactly what we get for $\Delta_s W$ if we put $l = 0$ in Eq. (208a). Hence, if we drop the restriction that

$l > 0$ in using the latter formula, we can forget the special spin correction entirely.

Upon adding $\Delta_l W$ and $\Delta_s W$ from (207) and (208a) for the n levels with $j = l + \frac{1}{2}$ and expressing the result in terms of j , we obtain the second term on the right in Eq. (206). The same result is obtained, in terms of j , for $j = l - \frac{1}{2}$, if we use (207) and (208b). Hence, the $n - 1$ levels with $j = l - \frac{1}{2}$ coincide with $n - 1$ of the levels that have $j = l + \frac{1}{2}$, and the number of distinct levels, as already stated, is only n .

The difference between the values of $\Delta_s W$ in (208a) and (208b) also gives the doublet spacing as stated in Eq. (203) above.

149. Fine Structure of Spectral Lines from One-electron Atoms. (a) *The Theoretical Patterns.*—The formulas for the energies fix the frequencies of the spectral lines to be expected. If the pair of levels having given values of n and j is treated as a single multiple level, no attention need be paid to the selection rule for l in determining the allowed jumps between such levels; it is sufficient to observe the selection rule for j (Sec. 140):

$$\Delta j = 0 \text{ or } \pm 1.$$

For, if a jump is allowed by this rule, component levels satisfying the condition that $\Delta l = \pm 1$ can always be selected.

As an example, the levels and allowed jumps for $n = 3$ and $n = 2$ are shown in Fig. 104. (The spacings between the j levels are shown to scale, but on the same scale the distance between the two groups would be some 24,000 times larger than as shown.) The selection rule for j allows five transitions, as shown by the arrows. When changes in l are considered as well, we find seven transitions, two pairs of them producing identical frequencies; denoting $l = 0, 1, 2$ by S, P, D as usual, we have as transitions: $3D_{5/2} \rightarrow 2P_{3/2}$, $3D_{3/2} \rightarrow 2P_{3/2}$, $3D_{3/2} \rightarrow 2P_{1/2}$, $3P_{3/2} \rightarrow 2S_{1/2}$, $3P_{1/2} \rightarrow 2S_{1/2}$, $3S_{1/2} \rightarrow 2P_{1/2}$, and $3S_{1/2} \rightarrow 2P_{3/2}$. The relative spacing of the five distinct lines is shown in the lower part of Fig. 104, the heights of the lines as drawn representing the theoretical relative intensities of the spectral lines on the assumption that all five quantum states for $n = 3$ are equally excited (i.e., that as many atoms are excited into one state as into another).

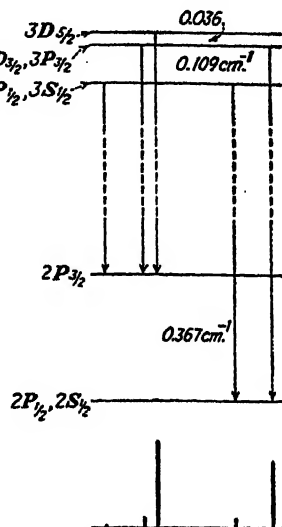


FIG. 104.—Theoretical fine structure of the $H\alpha$ line ($n = 3 \rightarrow n = 2$) for atomic hydrogen.

(b) *Comparison with Experiment for Hydrogen.*—The “lines” of the Balmer series of hydrogen were early observed as close doublets. In 1887, Michelson and Morley measured the doublet separation for the line of longest wave length, $H\alpha$ or $n = 3 \rightarrow n = 2$, and found for it 0.253 cm^{-1} . Subsequent measurements by other investigators gave values ranging from 0.293 to 0.357. The modern interpretation of the doublet appearance is that the five component lines are smeared together because of the Doppler broadening of all lines due to the thermal motion of the molecules, but two of them are much more intense than the others. The only means of comparing the observations with the theoretical predictions is, therefore, to calculate the contour of the total line from the theory, using the theoretical separations of the fine structure and the theoretical estimates of the relative

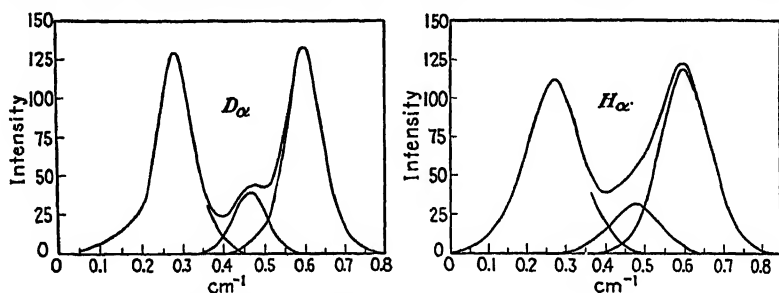


FIG. 105.—Typical intensity curves obtained for the $H\alpha$ line from hydrogen ($H\alpha$) and deuterium ($D\alpha$). Intensity is on an arbitrary scale, and the abscissa is numbered from an arbitrary point. (After Williams.)

intensities of the component lines, and making an approximate allowance for the Doppler effect.

The line $H\alpha$ has been studied very carefully in this manner. In Fig. 105 are shown typical curves obtained by R. C. Williams¹ for the spectral distribution of intensity in the $H\alpha$ line from ordinary hydrogen and from deuterium, or heavy hydrogen, the atoms of which are about twice as heavy as those of ordinary hydrogen and ought, therefore, to be influenced less by the Doppler effect [Sec. 164(a)]. The discharge tube was cooled by liquid air in order to minimize the broadening. The light from the tube was dispersed by a triple-prism spectrograph with a quartz Fabry-Perot etalon placed in the parallel beam of the collimator, and photographs were taken of the spectral region containing the $H\alpha$ line. The blackening on the film was measured with a Moll microphotometer using a thermocouple and galvanometer, and from these measurements the relative distribution of intensity in the line could be determined.

¹ WILLIAMS, *Phys. Rev.*, vol. 54, p. 558 (1938).

Only three of the theoretical component lines are evident in Fig. 105; one of the others is too close to the left-hand main line to be resolved, and the remaining one is evidently too weak to be seen. Theoretical line shapes, as determined by the Doppler effect, are drawn in the figure for the three lines, the assumed strengths of the lines being so adjusted as best to reproduce the observed curve of resultant intensity. The strengths of the lines as thus inferred agree only roughly with the theoretical predictions. The left-hand line ought to be relatively somewhat stronger and the middle one only about half as strong. Such deviations might be due, however, to unequal excitation of the initial levels ($S_{3/2}$, $P_{3/2}$, $D_{3/2}$). A much more serious discrepancy is that the *spacing* of the lines does not quite agree with the theoretical predictions. The distance between the two main peaks was found by Williams to be consistently 0.319 to 0.321 cm^{-1} for $D\alpha$, 0.315 to 0.319 cm^{-1} for $H\alpha$, whereas the theory predicts, from the level separations as shown in Fig. 104, $0.367 - 0.036 = 0.331$ cm^{-1} . The small central peak seems also to occur at about 0.134 cm^{-1} from the right-hand one, whereas the theory gives for this separation 0.109 cm^{-1} . The cause of these discrepancies between theory and observation—whether they are due to some unrecognized experimental complication or to some further slight correction required in the theory—is not yet known.¹

The remark may be added that, because of the fine structure and the relativity correction, the simple Balmer formula for the frequencies is not quite adequate. The following empirical formula of the more general Rydberg type gives very closely the wave numbers of the centers of the lines of the Balmer series of hydrogen when seen unresolved:²

$$\tilde{\nu} = 109,678.28 \left[\frac{1}{(2 - .00000383)^2} - \frac{1}{(m + .00000210)^2} \right],$$

$$m = 3, 4, 5, \dots$$

(c) *Ionized Helium*.—According to Eq. (206a), the fine-structure separations should be 2^4 or 16 times as great in lines emitted by an ionized helium atom as in the corresponding lines from hydrogen. A careful study was made by Paschen³ of the line $\lambda = 4,686$ Å from ionized helium. At the time, his observations were important because

¹ Cf. DRINKWATER, RICHARDSON, and W. E. WILLIAMS, *Roy. Soc., Proc.*, vol. 174, p. 164 (1940).

² FOWLER, "Report on Series in Line Spectra," Fleetway Press, London, 1922.

³ PASCHEN, *Ann. d. Physik*, vol. 82, p. 689 (1927).

they confirmed the wave-mechanical theory as against Sommerfeld's extension of Bohr's theory.

The line in question results from the jump $n = 4$ to $n = 3$ and so corresponds to the longest line of Paschen's infrared series for hydrogen. From the energy-level diagram, shown in Fig. 106, it is evident that there should be eight fine-structure lines, here labeled *abcdefgh*.

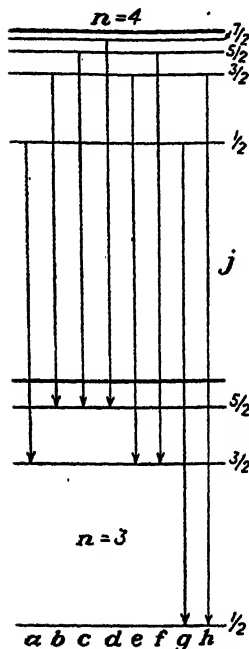


FIG. 106.—Diagram for the fine structure of the line $\lambda 4,686$ ($n = 4 \rightarrow n = 3$) of ionized helium. The two heavy lines indicate the levels according to the Bohr theory.

According to Sommerfeld's theory the lines *c* and *g*, arising from jumps in which $\Delta j = 0$, should not appear. In place of $j + \frac{1}{2}$ in the formulas, Sommerfeld wrote a quantum number referring to the orbital angular momentum, and this quantum number was subject to the requirement that in a transition it must change by unity. This rule rested on grounds equivalent to those that lead to the modern rule that $\Delta l = \pm 1$, and it seemed to be very hard to eliminate it from the theory.

Paschen's observations are compared with the theoretical predictions in Fig. 107. Components *a* and *b*, as predicted by both theories, are undoubtedly unresolved in Paschen's *ab*. The weak components, *c* and *e*, were probably unresolved from *d* and *f*. But Paschen found a fairly strong component *g*, which is predicted by the newer theory but is absent in the older one. This fact is decisive in favor of wave mechanics.

150. Two-electron Spectra.—The type of atom next simplest in its spectroscopic properties after those hitherto discussed consists of atoms which contain, when in their normal state, 2 spectroscopically active electrons outside of closed subshells.

In one subclass of such atoms, there are in the normal atom 2 valence electrons with *s* wave functions. These electrons themselves really form an additional closed subshell; but, when such an atom is excited by changing 1 valence electron to a wave function of higher energy, the subshell is no longer closed. It can also happen, as an alternative, that both valence electrons are thus excited simultaneously. Among neutral atoms of this type may be mentioned helium; then the alkaline earths, beryllium, magnesium, calcium, strontium, and barium; and the related group with larger cores formed by zinc, cadmium, and mercury. Ions with similar electronic exteriors

are C^{++} , Al^{+} , Si^{++} , Pb^{++} (the number of plus signs indicating the number of positive charges on the ion).

A second subclass of atoms with two optically active electrons is formed by those which, in their normal states, contain two s and two p valence electrons. In such cases the two s electrons usually (but not always) stay put, only the two p electrons being active. Examples of such atoms are neutral carbon, silicon, germanium, tin, and lead. Furthermore, certain observed spectra of the same type have been ascribed to singly (positively) ionized atoms of nitrogen, phosphorus, and bismuth.

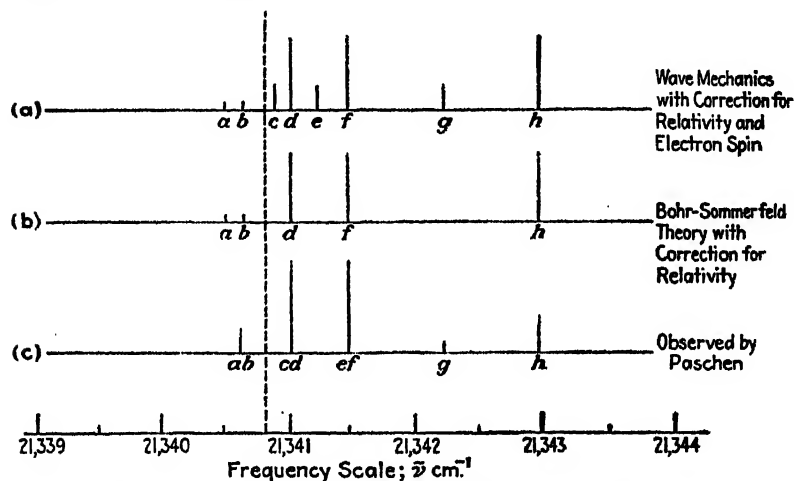


FIG. 107.—Fine structure of the helium line $\lambda 4,686$ (cf. Fig. 106). The dotted line gives the position of the single line predicted by Bohr's theory. Lengths of lines give relative intensities (qualitatively).

In order to arrive at a theoretical classification of the spectral terms to be expected for such atoms, let us first suppose the spin-orbit interaction to be omitted, so that we have the necessary conditions which make for LS coupling. Then, as stated in Sec. 141, it is possible, in spite of the electrostatic repulsions between the electrons, to fix their *total orbital angular momentum*. This is in accord with the classical analogy; for, in consequence of Newton's third law of motion, electrostatic interaction between parts of a system has no tendency to alter its total angular momentum. The quantum states for the atom will then be characterized by a quantum number L .

Under the same conditions, nothing prevents us from choosing the atomic quantum states so as to fix the *resultant spin angular momentum* as well, as represented by a quantum number S . The total number of electrons being even (as it is in any closed subshell),

S is restricted to integral values; actually, for a reason to be explained presently, it is restricted to the values 0 and 1.

Then, finally, resultant orbital and resultant spin angular momenta can be added into a *total resultant momentum*, as described in Sec. 144, with the introduction of a quantum number J . Terms with $S = 0$ will be singlet terms, having $J = L$. Terms with $S = 1$ will have a multiplicity $2S + 1 = 3$; they will contain three levels, for which $J = L + 1, L$ or $L - 1$, respectively, except that the S terms ($L = 0$) are, of course, single, with $J = S$. *Thus the LS terms for 2 electrons fall into two classes, singlet terms ($S = 0$) and triplet terms ($S = 3$).* They are denoted by such symbols as

$$^1S, ^3S; \quad ^1P, ^3P; \quad ^1D, ^3D;$$

and so on.

The differences in energy between these LS terms result from differences in the wave functions and in the effects of the various electrostatic interactions of the electrons with the nucleus and with each other. No splitting of these terms occurs so long as spin-orbit effects are ignored. When, however, we restore the spin-orbit terms in the wave equation, the various J levels composing each term become more or less separated from each other. Thus, for 2 electrons, we have finally the following array of J levels, in case LS coupling holds to a good approximation:

$$^1S_0; \quad ^3S_1; \quad ^1P_1; \quad ^3P_2, ^3P_1, ^3P_0; \quad ^1D_2; \quad ^3D_3, ^3D_2, ^3D_1; \dots$$

Perhaps it may seem surprising that singlet and triplet terms, differing apparently only in spin features, should be separated in energy by *electrostatic* interaction. The explanation is that the difference in S is necessarily associated with a difference in the *spatial* distribution of the wave function as well, and this latter difference results in a different electrostatic effect. The separation of singlet and triplet terms having the same L value, such as two P or two D terms, is caused by the exchange effect (Sec. 129).

So far we have said nothing about quantum numbers of the *individual electrons*. If we wish to introduce these into the picture, we must start a little further back in the perturbation-theory approach than we had to do in order to introduce the quantum numbers L and S . Let us begin by dropping *both the spin-orbit effect and the electrostatic interaction* of the 2 valence electrons with each other, except in so far as the latter is included in the central field (*cf.* Secs. 127 and 128). Then the 2 valence electrons will be represented by

two 1-electron wave functions with quantum numbers

$$n_1 l_1 \lambda_1 \mu_1; \quad n_2 l_2 \lambda_2 \mu_2.$$

As stated in Sec. 128, however, the electronic energies depend only on

$$n_1, l_1; \quad n_2, l_2.$$

A set of values of n 's and l 's is said to define a *configuration* of electrons. For example, the configuration $2s3d$ contains 1 electron with $n = 2$, $l = 0$, and another with $n = 3$, $l = 2$, values of λ and μ not being specified.

Let us fix our attention upon one such configuration. Out of the zero-order wave functions belonging to it, new wave functions can be formed representing states of the atom characterized by quantum numbers L and S . The method of doing this is similar to that by which orbital and spin momenta are added to form a resultant characterized by the quantum number J (Sec. 144). There is a variety of wave functions belonging to the given configuration, as indicated by different values of λ_1 , λ_2 , μ_1 and μ_2 . These functions are first combined in such a way as to obtain new functions representing states with definite values of the total orbital momentum, represented by the quantum number L . Such a function is obtained for each integral value of L from $l_1 + l_2$ down to $|l_1 - l_2|$:

$$|l_1 - l_2| \leq L \leq l_1 + l_2.$$

Similarly, we add the two electronic spin momenta. Since for each electron $s_1 = s_2 = \frac{1}{2}$ (Sec. 126), the maximum value of S is $s_1 + s_2 = 1$, and, since the values of S must be integrally spaced, the only other possible value is $S = 0$. In all of this work we may ignore the closed shells, since, as stated in Sec. 142, their contributions to the angular momenta are zero. The relations between the quantum numbers can be represented in the usual way by means of a vector diagram.

For example, as illustrated in Fig. (108a), the configuration $n_1 p, n_2 p$ (with $n_1 \neq n_2$ —see Sec. 152) gives rise to terms with

$$L = l_1 + l_2 = 1 + 1 = 2,$$

with $L = 1$, and $L = |l_1 - l_2| = 0$. (In the figure all L or S vectors are supposed drawn from the lowest point.) The configuration $n_1 p, n_2 d$ (with $l_1 = 1$, $l_2 = 2$, and $n_1 \neq n_2$) gives $L = 3, 2$, or 1 [Fig. 108(b)] The two ways of adding the spins are shown in Fig. 108(c). The student will find it instructive to check the following illustrations of the levels that result from certain other configurations of 2 electrons

($n_1 \neq n_2$):

$$\begin{aligned} n_1s, n_2s &\rightarrow {}^1S_0, {}^3S_1; & n_1s, n_2p &\rightarrow {}^1P_1, {}^3P_{0,1,2}; \\ n_1p, n_2p &\rightarrow {}^1S_0, {}^3S_1; & {}^1P_1, {}^3P_{2,1,0}; & {}^1D_2, {}^3D_{3,2,1}. \end{aligned}$$

It is customary to add the specification of the configuration, when this is of interest, in the spectroscopic designation of the term. Thus, we find in the literature such symbols as

$$3s4d {}^1D_2; \quad 2s^2 {}^1S_0.$$

(Here $2s^2$ means $2s2s$, *i.e.*, 2 electrons with $n = 2$, $l = 0$.)

The principal advantage in specifying an electronic configuration lies in the existence of a *selection rule for configurations*. The general

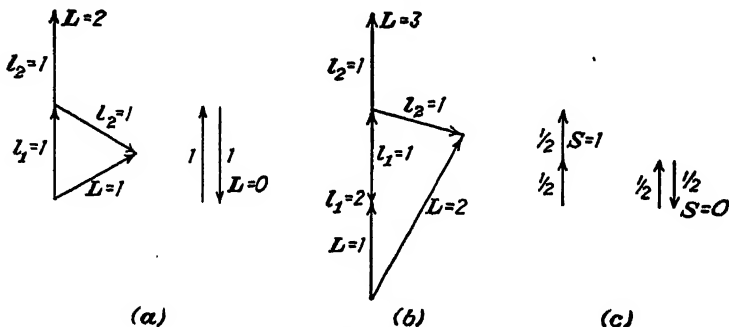


FIG. 108.

rule stated in Sec. 138, that only 1 electron jumps at a time, requires that *radiative transitions occur, in general, only between levels belonging to two configurations which differ in just one electronic wave function. Furthermore, the difference in the two values of l for this 1 electron must be exactly unity.* Thus transitions may occur between a level belonging to a $2s3p$ configuration and one belonging to a $2s3d$ configuration, but not between $2s3p$ and $2p3d$, because here *both* electrons have different wave functions, and not between $2s3p$ and $2s4f$, because here $\Delta l = 2$ for 1 electron. This selection rule for configurations effects a great simplification, because it diminishes the number of possible transitions that have to be considered.

In the stage of perturbation theory which we are considering at the moment, all terms belonging to a given configuration have the same energy. If we then restore the complete electrostatic interaction between the electrons in the wave equation, continuing to omit only the spin-orbit interaction, the terms for different values of L and S become separated. The quantum numbers L , S , and J still retain their full

significance. Thus we arrive at the stage of perturbation theory first contemplated in the present section. The significance of the individual electronic quantum numbers is more or less impaired, however, by the introduction of the full electrostatic interaction. In this circumstance lies the reason for describing the application of *LS* coupling before mentioning electronic configurations.

To form wave functions for the atom with full allowance for the electrostatic interaction, it is necessary to combine, to a greater or less extent, functions belonging to different configurations. Often, however, one configuration predominates in each combination; then the selection rules for configurations hold approximately, and it is convenient to retain the symbol for the configuration as part of the label for the atomic level. In some cases, more especially when 3 or more electrons are involved, the selection rules for configurations fail so badly that the assignment of values of n and l to the individual electrons becomes quite meaningless.

151. The Spectrum of Mercury.—As an interesting example of 2-electron spectra, the familiar arc spectrum of mercury may be selected. Its chief energy levels and many lines are shown in Fig. 109. The levels that are considered to form a series are placed under each other, with an appropriate spectroscopic symbol at the top. The electronic configuration for the 2 valence electrons is stated completely for the lowest levels, and with the omission of one 6s electron for a few others. Many of the chief spectral lines are shown by lines connecting the levels, wave lengths being given in angstroms; heavy lines on the diagram indicate spectral lines that are relatively strong in the arc spectrum.

A glance at the figure shows that the levels fall naturally into two classes, singlet and triplet levels. A number of intercombination lines occur, however, showing that *LS* coupling is not complete in mercury. The strongest of these intercombination lines, $6s6p\ 3P_1 \rightarrow 6s^2\ ^1S_0$, is one of the principal ultraviolet lines in the mercury spectrum. Its strength, however, is not necessarily an indication of a high transition probability between these levels, owing to a high value of Einstein's A (Sec. 125). It is more likely that the strength of this line is due to a piling up of atoms in the $6s6p\ ^1P_1$ level. As they are thrown into this state as a result of collisions with electrons in the arc or drop into it from higher levels, they have nowhere else to go, therefore they accumulate until through sheer numbers they are able to leave as fast as they enter. It is somewhat like water piling up in a pool until its depth becomes great enough to force a large discharge from the pool through a small drain.

Contrary to the selection rule for S , that for L is pretty well obeyed. With values of L assigned as in the diagram, S terms combine only with P terms, P with S and D , and so on. However, a few weak lines, not shown on the diagram, have been observed corresponding to $\Delta L = 2$.

All of the transitions shown in Fig. 109 are in harmony with the selection rule for configurations. Only 1 electron changes its n and l , and always $\Delta l = \pm 1$.

The spectral lines can be grouped into *series*, if desired. Thus, within the singlet system, all lines ending on the lowest 1S_0 level form the singlet principal series. Only these and the line $\lambda = 2,536 \text{ \AA}$ can be observed in absorption in mercury vapor. Of the lines ending on the lowest 1P level, those originating from 1S terms form a sharp series, those originating from 1D terms, a diffuse series, just as in sodium; and so on. Similar series can be picked out within the triplet system. It is really not very interesting to group the lines of such a complex spectrum into series, however, especially when the "fine structure" is as coarse as it is in the mercury spectrum. Thus the great spectral sextet of ultraviolet lines, $\lambda = 2,967$ to $\lambda = 3,663$, from the lowest 3D to the lowest 3P term, would constitute together the first "line" of the triplet diffuse series.

The student will recognize the lines $^1D_2 \rightarrow ^1P_1$ ($\lambda 5,791$), $^3S_1 \rightarrow ^3P_2$ ($\lambda 5,461$), and $^3S_1 \rightarrow ^3P_1$ ($\lambda 4,358$) as the familiar yellow, green, and blue lines emitted from the mercury arc.

The normal state of the mercury atom, as shown in the diagram, is a 1S_0 state arising out of the $6s^2$ configuration of the valence electrons. All of the levels shown on the diagram belong to configurations in which just 1 electron is excited from a $6s$ state into a different state. A few levels ascribed to the $6p^2$ configuration have also been discovered.

152. Equivalent Electrons.—In the configurations discussed in Sec. 150, n_1 and n_2 were assumed to be different. When 2 electrons have the same n and also the same l , they are called by spectroscopists *equivalent electrons*. In configurations containing equivalent electrons, such as $5s^2$ or $5s7p^3$, certain terms that would otherwise occur are excluded through the operation of the exclusion principle (Sec. 128).

Consider, for example, the configuration $1s^2$ in helium, or $6s^2$ in mercury. To form a wave function for such electrons in the zero-order stage of perturbation theory, we first multiply together two 1-electron wave functions each of which is characterized by suitable values of $n l \lambda \mu$. This gives us a product function, $\psi_{n_1 l_1 \lambda_1 \mu_1} \psi_{n_2 l_2 \lambda_2 \mu_2}$. In the configurations under discussion, $\lambda_1 = l_1 = 0$, $\lambda_2 = l_2 = 0$, but each μ may be $+\frac{1}{2}$ or $-\frac{1}{2}$. Thus four different product functions are

possible, which may conveniently be indicated by means of the quantum numbers alone, thus:

$$(n00\frac{1}{2}, n00\frac{1}{2}), \quad (n00\frac{1}{2}, n00 - \frac{1}{2}), \quad (n00 - \frac{1}{2}, n00\frac{1}{2}), \\ (n00 - \frac{1}{2}, n00 - \frac{1}{2}).$$

Of these four product functions, however, the first and last put both electrons into the same 1-electron state; hence, these are ruled out by the exclusion principle. The middle two differ only in that in the second one the 2 electrons change places in their wave functions, hence these two product functions give rise to the same wave function for the atom; for, as we have seen (Sec. 128), it makes no difference which electron is assigned to which state, or, in other words, each combination of $n l \lambda \mu$'s, regardless of their order, gives just one quantum state for the atom. Thus, we obtain out of the configuration just one wave function. This must be the single function for a 1S_0 state, with $L = 0$, $S = 0$, $M = J = 0$. The 3S_1 state, therefore, which would require three additional functions, with $M = 1, 0$, or -1 , is missing.

This is the reason that no 3S_1 state, arising from the $6s^2$ configuration for mercury, is shown in Fig. 109. The normal state, 1S_0 , stands alone. The same happens to be true of the normal state for all atoms with just 2 valence electrons.

General rules can be given for determining which states are allowed for any given configuration containing identical electrons.¹ All of the J levels belonging to a given LS term appear or drop out together. Thus, the rule stated above for ascertaining the possible J values to go with given L and S is perfectly general; the restriction in the case of identical electrons applies only to the rule for determining the possible L and S values that can be formed with given l 's and s 's.

153. "jj" Coupling.—The validity of LS coupling is limited to atoms that are not too heavy. As the atomic number increases, the spin-orbit effects become rapidly larger; as a consequence, the J levels tend less and less to group themselves into LS multiplets, and the selection rules for L and S fail more and more. Finally, in very heavy atoms the spin-orbit effects tend to predominate over the special electrostatic effect (*i.e.*, over the part of the electrostatic effect that is not included in the central field). Then an approximation to another type of "coupling," known as " jj coupling," occurs.

This term has reference to an alternative path of approach in applying the perturbation theory. Suppose, instead of omitting the spin-orbit terms from the wave equation, we first omit the terms representing the effect of electronic repulsion, except in so far as the

¹ Cf. WHITE, "Introduction to Atomic Spectra," 1934.

effect is allowed for by the departure of the central field from the nuclear field. Then we have a set of electronic wave functions very similar to those described for the 1-electron atom in Sec. 128. These electronic wave functions, and the associated energies, are determined by the central field and are characterized by quantum numbers $n\ l\ j\ m$. A state of the whole atom is characterized, in zero order, by a set of these quantum numbers for each of the N electrons in the atom:

$$n_1 l_1 j_1 m_1, \quad n_2 l_2 j_2 m_2, \quad \dots, \quad n_N l_N j_N m_N. \quad (B)$$

These zero-order atomic states now take the place of those that were previously constructed out of $n\ l\ \lambda\ \mu$ functions; and expression (B) takes the place of (A) in Sec. 128. Here each electron has fixed values of resultant angular momentum, orbital plus spin; the magnitude of the resultant momentum is $\sqrt{j(j+1)}\ h/2\pi$, and its component in the direction of some chosen axis is $m h/2\pi$. Furthermore, we are now including the spin-orbit terms in the wave equation. Consequently the energy of the atom, when in a zero-order state as specified by (B), depends upon the values of the electronic quantum numbers

$$n_1 l_1 j_1, \quad n_2 l_2 j_2, \quad \dots, \quad n_N l_N j_N.$$

Thus, if we consider only states belonging to a particular configuration specified by the quantum numbers

$$n_1 l_1, \quad n_2 l_2, \quad \dots, \quad n_N l_N,$$

they fall into several j -groups characterized by the various combinations of the j 's that can occur in the presence of the given values of $l_1 l_2, \dots, l_N$. These groups correspond to the LS terms in LS coupling. Their number cannot exceed 2^N ; for each

$$j_\tau = l_\tau \pm \frac{1}{2}, \quad (\tau = 1, \dots, N),$$

except that, if $l_\tau = 0$, $j_\tau = l_\tau + \frac{1}{2} = \frac{1}{2}$.

Between the j -groups, there will be the usual selection rules referring to the electronic wave functions:

1. Only one $n\ l\ j$ set of quantum numbers can change in a radiative transition—"only 1 electron jumps at a time."
2. For the jumping electron,

$$\Delta l = \pm 1, \quad \Delta j = 0 \text{ or } \pm 1.$$

The wave functions in a particular j -group can then be combined to form new wave functions that fix the total angular momentum and are characterized, as usual, by quantum numbers J and M (Sec. 140).

At this stage of perturbation theory all J levels belonging to the same j -group have the same energy. When the finer effects of electrostatic interaction are then included, the J levels of each j -group separate slightly from each other and form a sort of jj multiplet.

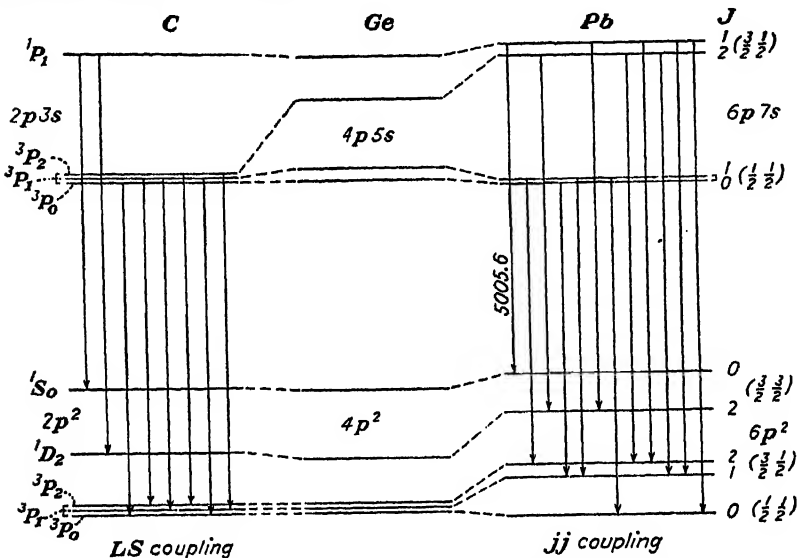


FIG. 110.—A transition from LS to jj coupling. Within each multiplet of levels the spacing is to scale, except that levels joined by a brace should be drawn closer together. The ranges covered by the levels, in cm^{-1} , are: carbon, $2p^2$, 69,231.1 to 90,878.3; $2p3s$, 28,898 to 30,547; germanium, total range (term values uncertain), $4p^2$, 16,367; $4p5s$, 2,568; lead, $6p^2$, 30,365 to 59,821; $6p7s$, 10,383 to 24,863. One wave length is given in angstroms; the other lines lie in the ultraviolet.

Thus analogous designations of a J level with $J = 3$, in LS and in jj coupling, might be

$$\begin{array}{ccc}
 LS & & jj \\
 n_1 l_1 \cdots n_N l_N {}^5D_3; & (n_1 l_1 \cdots n_N l_N)(j_1 \cdots j_N)_3.
 \end{array}$$

When equivalent electrons occur in a configuration, certain jj groups are absent, just as certain LS terms are absent in the case of LS coupling.

An interesting example exhibiting the transition from LS to jj coupling is shown in Fig. 110. The relative positions of certain levels are shown for carbon ($Z = 6$), germanium (32), and lead (82), corresponding levels being connected by dotted lines; and all transitions that are observed to occur between these levels in carbon and in lead are indicated by arrows. In silicon (14), the corresponding set of levels is observed to be arranged much as in carbon, whereas tin (50) resembles lead in this respect.

In Fig. 110, we note that the J value of any level remains the same in all three spectra; but in carbon the levels group themselves by their energy values into good LS multiplets, whereas in lead they form jj groups. The radiative transitions indicated in the figure for carbon and for lead, as observed, include all that are allowed by the selection rules in the two cases. The student will find it instructive to verify this statement in detail and to note which lines occur with one form of coupling and not with the other.

Many other examples of the same sort could be given.

154. Spectra Due to Three or More Electrons.—When more than 2 electrons are optically active, the situation is naturally more complicated, but it presents no radically new features. If the spin-orbit effects are relatively weak, so that LS coupling is possible, we first add the orbital angular momenta of all the electrons. The largest value of L that can occur in a given configuration containing no identical electrons equals the sum of the l 's for the individual electrons; all smaller integral values of L then occur that are consistent with vector addition of the l -vectors. Similarly, the electronic spins are added into a resultant spin vector. The largest possible value of S for N electrons, when no two are equivalent, is

$$s_1 + s_2 + \cdots + s_N = \frac{N}{2},$$

and all positive values occur that are integrally spaced downward from this maximum value. If the configuration contains equivalent electrons, however, certain of the LS terms deduced in this way are absent.

Evidently if N is even, S must be integral, with possible values 0, 1, 2, \cdots , $N/2$; but if N is odd, S is half-integral, with values $\frac{1}{2}$, $\frac{3}{2}$, \cdots , $N/2$. Thus, as the number of active electrons increases, there is an alternation of multiplicities, as well as an increase in their variety. A single active electron leads to doublet levels; 2 electrons to singlets and triplets; 3, to doublets and quartets ($S = \frac{1}{2}$ or $\frac{3}{2}$); 4, to singlets, triplets, and quintets ($S = 0, 1, 2$); and so on. Even octet terms have been observed (*e.g.*, in the arc spectrum of manganese).

In certain configurations containing equivalent electrons, certain kinds of multiplicity are missing, just as we found 2 ns electrons to give rise to a singlet term but not to a triplet. The extreme example of this is presented by any closed subshell, of which ns^2 is a special case. When so many electrons are present with given n and l that in writing wave functions for them every allowed value of λ and μ must be used, so as to avoid conflict with Pauli's exclusion principle,

then there is just one way of selecting a set of 1-electron states for the electrons in the subshell (Sec. 128). Thus only a single state for the atom is obtained. The value of J must then be zero, for any other value requires the existence of several wave functions with different values of M . Actually LS coupling holds in such cases, so that a 1S_0 state results. Thus, any closed subshell by itself gives rise to a 1S_0 level; and the result of combining any number of closed subshells is also a 1S_0 level. The normal state of every inert gas, therefore, is a 1S_0 state.

The spectroscopic character of the normal state for an isolated atom of each of the elements is given in the table in Appendix III. For further information concerning complex spectra, the student is referred to special treatises on the subject.¹

155. The Effect of a Magnetic Field on an Atom.—When an atom is subjected to the action of *external forces*, its quantum states are modified. States which were formerly degenerate, corresponding to the same energy, may be separated, so that the degeneracy is reduced. The spectral lines emitted by the atom may consequently be altered or even split into several components.

The most interesting case of this sort is the *Zeeman effect*. In Sec. 43, Chap. III, we discussed it in terms of classical theory, and we remarked that in some cases the predictions of classical theory are confirmed by experiment. Such cases are sometimes referred to as the *normal Zeeman effect*. More commonly, however, the observed pattern of lines is quite different from that predicted by classical theory; this is called the *anomalous Zeeman effect*. No explanation of the anomalous effect was found until electron spin was introduced. Furthermore, in the modern wave-mechanical treatment of the Zeeman effect, there is, naturally, nothing closely resembling the electronic motion as described in the classical theory. Nevertheless, certain features of the classical picture still retain an interest, either because they are actually preserved in the new theory, or because they furnish a useful picture in terms of which the phenomenon can readily be visualized.

In the new theory, as in the old, the changes produced in the energy levels by a magnetic field can be regarded as resulting from the possession by the atom of a *magnetic moment*. Classical theory indicates that there should be a close relation between the magnetic moment due to motion of an electron in an orbit and the angular

¹ WHITE, *op. cit.*; HERZBERG, G., "Atomic Spectra and Atomic Structure," 1937.

momentum resulting from the same motion. It will be instructive to work out this relation.

Suppose a particle carrying algebraic charge q revolves in a fixed circular orbit of radius a at a velocity v (Fig. 111). Then it makes $v/(2\pi a)$ revolutions per second and so is equivalent to a current I flowing around the same circle of magnitude

$$I = \frac{qv}{2\pi a}.$$

It is well known that a current I , flowing in any plane loop enclosing area A , is equivalent to a magnet placed with its axis perpendicular to the loop and having a magnetic moment

$$\mu = \frac{IA}{c}$$

(q and I being in electrostatic units, hence the factor $1/c$). In the present case $A = \pi a^2$, therefore,

$$\mu = \frac{I\pi a^2}{c} = \frac{qva}{2c}.$$

The moment μ can be treated as a vector perpendicular to the plane of the orbit; let us write for it \mathbf{u}_l .

The particle will also have angular momentum about the center of the circle of magnitude

$$G_l = mva,$$

m being its mass. As a vector, the angular momentum is likewise perpendicular to the orbit.

Comparing these two equations, we see that the magnetic moment and angular momentum are related by the vector equation

$$\mathbf{u}_l = \frac{qG_l}{2mc}. \quad (209a)$$

This result, which is independent of a , can be shown to hold for motion in an orbit of any shape. If the particle is an *electron* with numerical charge e , we have $q = -e$ and

$$\mathbf{u}_l = -\frac{eG_l}{2mc}. \quad (209b)$$

Thus, for an electron, \mathbf{u}_l and G_l are oppositely directed.

A similar expression must then be added for the effect of *electron spin*. According to classical electromagnetic theory, a sphere, uni-

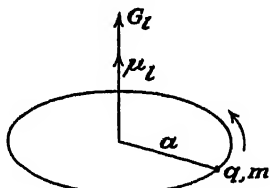


FIG. 111.—Diagram illustrating the angular momentum G_l and magnetic moment μ_l due to a charge revolving in a circle.

formly charged with electricity on its surface and set spinning, behaves as if it were covered with electric currents and possessed a corresponding magnetic moment. Because of the momentum in the electromagnetic field around it, it also possesses angular momentum. For the relation between its vector magnetic moment \mathbf{u}_s and its vector angular momentum \mathbf{G}_s classical theory gives

$$\mathbf{u}_s = \frac{q\mathbf{G}_s}{mc}, \quad (210a)$$

where q/c is the algebraic charge on the sphere in electromagnetic units and m is its electromagnetic rest mass, due to the inertia of its electric field (*cf.* Sec. 39). For an electron this becomes

$$\mathbf{u}_s = -\frac{e}{mc}\mathbf{G}_s. \quad (210b)$$

Comparing Eqs. (209b) and (210b), we note that twice as great a magnetic moment is associated with angular momentum of spin as with orbital angular momentum.

Now, there is no reason to suppose that \mathbf{u}_l and \mathbf{u}_s would necessarily have fixed directions for an electron in an atom. If a magnetic field is present, however, the components of the orbital and spin momenta in the direction of the field will have certain average values, which we may denote by \bar{G}_{lH} and \bar{G}_{sH} , respectively. According to Eqs. (209b) and (210b), this average momentum will be accompanied by an average magnetic moment in the direction of the field of magnitude

$$\bar{\mu} = -\frac{e}{2mc}(\bar{G}_{lH} + 2\bar{G}_{sH}) \quad (211)$$

(in which e stands for the *numerical* electronic charge). Or, if the atom contains more than 1 electron, its magnetic moment in the direction of the field will be

$$\bar{\mu} = -\frac{e}{2mc} \sum (\bar{G}_{lH} + 2\bar{G}_{sH}), \quad (211a)$$

the Σ indicating a summation over all of the electrons in the atom.

These same expressions for $\bar{\mu}$ are given by wave mechanics. The occurrence of a magnetic moment associated with spin follows from the Dirac relativistic equation in much the same way as does the angular momentum of spin (Sec. 126). The quantum-mechanical definitions of the averages are

$$\bar{G}_{lH} = \int \psi^* G_{lH} \psi dq, \quad \bar{G}_{sH} = \int \psi^* G_{sH} \psi dq,$$

G_{lH} and G_{sH} in the integrals standing for operators and ψ^* for the com-

plex conjugate of the wave function ψ ; the integrations are to be extended over all values of the coordinates of all electrons in the atom.

In passing, it may be remarked that in Eq. (211a) we have the wave-mechanical basis for the treatment of paramagnetism and diamagnetism. *Paramagnetism* is ascribed to a tendency of an atom to occur more frequently in those quantum states in which it has a magnetic moment in the direction of the magnetic field than in states in which the moment is opposed to the field. *Diamagnetism* results from a slight modification of the properties of the quantum states themselves, by which a slight additional magnetic moment is produced, always in a direction opposed to the field, in accordance with Lenz's law. *Ferromagnetism* is a more complicated phenomenon, believed to be due to a lining up of electronic spins in different atoms under special conditions.

For our purposes, however, we are interested chiefly in effects upon the *atomic energy levels*. In classical theory, a magnet of moment M in a uniform magnetic field H behaves as if it possessed a potential energy of magnitude

$$-MH \cos (M, H),$$

the symbol (M, H) standing for the angle between the directions of M and H . Here $M \cos (M, H)$ can also be regarded as representing the component of the moment in the direction of the field. Similarly, wave mechanics indicates that the effect of the magnetic field upon an atomic energy level is to change the energy by the amount

$$W_H = -\bar{\mu}H = \frac{e}{2mc} \sum (\bar{G}_{Hn} + 2\bar{G}_{\perp n}) H, \quad (212)$$

by Eq. (211a). If the field H is not too strong, a sufficiently accurate value of W_H can be obtained by using the values of \bar{G}_{Hn} and $\bar{G}_{\perp n}$ that belong to the various quantum states in the absence of a field.

The ideas and equations thus developed are adequate for an understanding and visualization of the theory of the Zeeman effect. It may be instructive, however, to investigate a little further the nature of the magnetic energy W_H due to the classical motion of an electron in an orbit. The student who prefers to do so can omit the next section entirely without embarrassment in reading the rest of the book.

156. Classical Theory of the Magnetic Energy.—The change in the energy of a revolving electron due to the presence of a magnetic field cannot be of the nature of an ordinary *potential* energy, for the force exerted by the magnetic field upon the electron acts always

perpendicularly to its velocity and so never does any work upon it. If, therefore, owing to some change in conditions, the energy decreases, this decrease must occur in the kinetic energy of the electron. (Actually there is also a reaction on the magnetizing system itself, but further study shows that this can be left out of consideration.)

To take a definite case, suppose an electron is revolving in a circular orbit of radius a , under the action of a force directed toward a fixed center (*cf.* Fig. 111). Let a magnetic field of H oersteds, perpendicular to the plane of the orbit, be created very slowly. If there were a conducting wire lying along the orbit, there would be an induced electromotive force in this wire of magnitude

$$\epsilon = -\pi a^2 \frac{dH}{dt}.$$

According to the elementary definition of e.m.f., this means that work of amount $-\epsilon e/c$ will be done on the electron every time it goes round the orbit, $-e/c$ being its charge in electromagnetic units. This work must be done by a force acting on the electron. Hence, if \bar{F} is the average component of this force in the direction of the motion, we have for the work done in a revolution

$$\begin{aligned} \bar{F} \times 2\pi a &= -\frac{\epsilon e}{c} = \frac{e}{c} \pi a^2 \frac{dH}{dt}. \\ \therefore \bar{F} &= \frac{ea}{2c} \frac{dH}{dt}. \end{aligned}$$

The average moment of force about the center, on the other hand, is $a\bar{F}$; and this must be equal to the rate of change of the angular momentum G_{IH} about an axis parallel to the field. Thus

$$\frac{dG_{IH}}{dt} = \frac{ea^2}{2c} \frac{dH}{dt};$$

and, integrating, we have for the whole angular momentum imparted to the electron during the creation of the field

$$\Delta G_{IH} = \frac{ea^2}{2c} H. \quad (213)$$

Two cases may now be distinguished.

If in the beginning the electron had no angular momentum, and hence also no orbital magnetic moment, it acquires angular momentum of magnitude ΔG_{IH} , as given by Eq. (213), and also, by Eq. (209*b*), a magnetic moment

$$\Delta\mu_H = -\frac{e}{2mc} \Delta G_{IH} = -\frac{e^2 a^2}{4mc^2} H.$$

A similar expression is obtained from wave mechanics for the magnetic moment developed in an atom, when it is in a given quantum state, by the application of a magnetic field:

$$\Delta\mu_H = -\frac{e^2 H}{4mc^2} \sum R^2, \quad (214)$$

where R^2 is the average of the square of the distance of any electron from an axis drawn through the nucleus parallel to the field, calculated as are the quantities \bar{G}_{iH} and \bar{G}_{eH} that occur in Eq. (211), and Σ indicates summation for all electrons in the atom. This added magnetic moment, necessarily opposite in direction to the field, gives rise to the phenomenon of diamagnetism.

The electron is also given a certain amount of kinetic energy, but this is proportional to H^2 and hence is easily shown to be negligible.

If, on the other hand, the electron is already revolving, then, although the same changes in momentum and moment occur, they are in actual cases small as compared with the initial values. Now, however, a change in the *kinetic energy* occurs that is of the *first order* in H . Let us denote the angular momentum about the direction of H by G_i , taking it to be positive, as is ΔG_i , when the associated rotation is clockwise to an observer looking in the direction of the magnetic field. Then, since $|G_i| = mva$, the total kinetic energy of the electron is

$$K = \frac{1}{2}mv^2 = \frac{G_{iH}^2}{2ma^2}.$$

Thus, when the angular momentum is changed from G_{iH} to $G_{iH} + \Delta G_{iH}$, G_{iH} now standing for the initial momentum, the kinetic energy is increased by

$$\Delta K = \frac{1}{2ma^2}[(G_{iH} + \Delta G_{iH})^2 - G_{iH}^2] = \frac{1}{ma^2}\left[G_{iH}\Delta G_{iH} + \frac{1}{2}(\Delta G_{iH})^2\right].$$

Here the last term is of the second order in H and is negligible. Hence, using (213), to the first order in H

$$\Delta K = \frac{e}{2mc} G_{iH} H.$$

ΔK may be positive or negative according to the sign of G_{iH} .

Comparing the value just found for ΔK with the first term on the right in Eq. (212) with the Σ omitted (for one electron), which represents the change in energy due to the field and may be denoted by W_{iH} , we see that $W_{iH} = \Delta K$. Thus W_{iH} is accounted for in classical

theory by the change that occurs in the kinetic energy of the revolving electron during the introduction of the magnetic field.

In obtaining this result, we supposed the plane of the orbit to remain fixed in position while the field varied. The same results are obtained, however, if, instead of creating the field, we rotate the plane of the orbit from a position parallel to the field into a position perpendicular to it. The revolving electron behaves just like a gyrostat mounted in frictionless bearings. Furthermore, the results are readily shown to hold for an orbit of any shape.

The method just described for finding W_{IH} was especially advantageous in developing the older quantum theory of the Zeeman effect, because it obviated all doubt as to the proper conditions for determining the allowed quantum states in the presence of a magnetic field. For it was accepted as a general principle that, if a system is in a certain quantum state to begin with and if the external conditions are changed very slowly, the system must remain in an allowed quantum state; whatever state it passes into must, therefore, be an allowed state under the new conditions. This principle, known as Ehrenfest's *adiabatic principle*, holds under wave mechanics as well.

The quantum states thus inferred for a 1-electron atom subject to a magnetic field are more simply described, however, from a distinctly different point of view, which lies much closer to the classical picture of the Zeeman effect. For this purpose we make use of a celebrated theorem due to Larmor, which for our purposes may be stated in the following form: Given a system of electrons describing orbits, however complicated, around a nucleus located at the origin of a set of coordinate axes to which the motion is referred. The behavior of the system when subjected to a magnetic field H may be described by saying that, neglecting small quantities of second order and higher, the electrons retain their original motions with respect to the coordinates, but that by the action of the field the whole system, electrons and coordinate axes, is given a rotation about an axis parallel to the field and passing through the nucleus, the angular velocity of the rotation being¹

$$\omega_L = \frac{eH}{2mc} \quad (215)$$

(e = numerical charge and m = mass of an electron, c = speed of light). This rotation is called the *Larmor precession*.

¹LARMOR, "Æther and Matter," p. 341, Cambridge University Press, London, 1900. See also RUARK and UREY, "Atoms, Molecules and Quanta," p. 759; JEANS, "Electricity and Magnetism," p. 565.

In view of this theorem we should expect to find the electron executing the same Bohr orbits relative to a frame of reference undergoing the Larmor precession as it does relative to a fixed frame in the absence of the field. It is easily shown that this assumption leads to the correct value for the energy.

The Larmor precession has no effect upon the potential energy of the electron. To find the resulting change in the kinetic energy, let us resolve the velocity of the electron at any instant into three rectangular components, of which one, v_1 , is perpendicular both to the magnetic field and to the radius drawn from the nucleus to the electron. Only this component of the velocity is affected by the precession; and it is only this component that contributes to G_{1H} , the angular momentum about the direction of the field. The precession adds to v_1 a component of velocity $R\omega_L$, R being the perpendicular distance of the electron from a line drawn through the nucleus parallel to H . The change in kinetic energy is, therefore,

$$\Delta K = \frac{1}{2}m[(v_1 + R\omega_L)^2 - v_1^2] = m(v_1R\omega_L + \frac{1}{2}R^2\omega_L^2).$$

But

$$G_{1H} = mv_1R, \quad \therefore \Delta G_{1H} = mR^2\omega_L.$$

Hence,

$$\Delta K = \frac{1}{mR^2} \left[G_{1H} \Delta G_{1H} + \frac{1}{2}(\Delta G_{1H})^2 \right].$$

This is obviously a generalized form of the first expression obtained above for ΔK . Therefore, our assumption concerning the Bohr orbits was justified.

We are now fully prepared for a discussion of the Zeeman effect on the basis of quantum theory.

157. Zeeman Effect in a Huge Field.—The simplest type of Zeeman effect should be produced in a field so strong that complications due to all other sources can be ignored. Such a field we shall call "huge."

Consider first 1 electron in a central field, of Coulomb type or otherwise, upon which is superposed a uniform magnetic field H . Let us take an axis parallel to the magnetic field in constructing wave functions for it characterized by the quantum numbers n, l, λ, μ (Sec. 128). When the electron is represented by one of these wave functions, its component of orbital angular momentum about the axis, G_{1H} , and its component of spin momentum, G_{sH} , have respective magnitudes

$$G_{1H} = \frac{\lambda h}{2\pi}, \quad G_{sH} = \frac{\mu h}{2\pi}.$$

Substitution of these values for \bar{G}_{1H} and \bar{G}_{sH} in Eq. (212), in which we

The cause of the separation of the levels can be visualized, if desired, with the help of a vector diagram similar to that used for the angular momentum. Let a vector of length l be drawn from a fixed point in such a direction that its component along a line parallel to the field is λ ; from another fixed point let another vector of unit length be drawn either parallel to the field or opposed to it (Fig. 113). Each pair of positions of these vectors then represents one of the quantum states characterized by λ and μ . The vectors can be regarded as representing magnetic moments due to orbital motion and to spin, of respective magnitudes l and 1 Bohr magnetons,¹ the directions of these moments being opposite to those of the vectors because of the negative charge on the electron.

Consider now *transitions* between two such sets of energy levels. We may write Eq. (216) for them thus:

$$W' = W'_0 + \frac{eh}{4\pi mc}(\lambda' + 2\mu')H,$$

$$W'' = W''_0 + \frac{eh}{4\pi mc}(\lambda'' + 2\mu'')H.$$

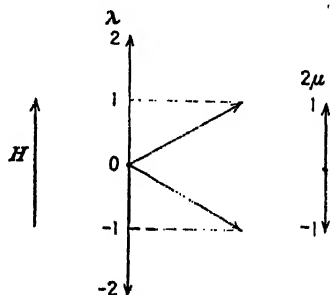


FIG 113.

As selection rules we have, besides $\Delta l = \pm 1$,

$$\Delta \lambda = 0 \text{ or } \pm 1, \quad \Delta \mu = 0.$$

For comparison with other cases we may note that these rules require also that

$$\Delta m = 0 \text{ or } \pm 1, \quad (218)$$

where $m = \lambda + \mu$ and represents, in units of $h/2\pi$, the total component of angular momentum in the direction of the field.

Calculating frequencies as $(W' - W'')/h$ and writing

$$\frac{(W'_0 - W''_0)}{h} = \nu_0$$

for the frequency of the line emitted in zero field, we find just three possible lines when the field is present:

$$\lambda'' - \lambda' = \Delta \lambda = 0: \quad \nu = \nu_0 \quad (\pi)$$

$$\Delta \lambda = -1: \quad \nu = \nu_0 + \frac{e}{4\pi mc}H \quad (\sigma)$$

$$\Delta \lambda = 1: \quad \nu = \nu_0 - \frac{e}{4\pi mc}H \quad (\sigma)$$

¹ Perhaps the vectors might better be drawn of magnitudes $\sqrt{l(l+1)}$, $\sqrt{1(1+1)}$, but this refinement is of no interest.

Each of these lines results, in general, from several different jumps, so that it might be described as consisting of several lines superposed. An example is illustrated in Fig. 112, already mentioned. The student will find it instructive to construct a similar diagram for transitions between levels with $l = 2$ and $l = 1$.

Comparing these results with those obtained from classical theory in Sec. 43, we see that we have here exactly the *classical triplet*. According to further conclusions from wave mechanics, even the polarization features should be the same as those deduced from classical theory. The undisplaced line, corresponding to $\Delta m = 0$, should be polarized with the electric vector in a plane parallel to the magnetic field, and it should be invisible when the radiation is viewed longitudinally (i.e., in a direction parallel to the field); such lines are called π lines (i.e., "parallel" as to polarization) or, sometimes, p lines. The other two lines, when the radiation is viewed transversely, should be plane polarized with the electric vector perpendicular to the magnetic field, whereas, when viewed longitudinally, they should be circularly polarized with opposite directions of rotation, as in the classical case; such lines are called σ lines (from the German "senkrecht" = perpendicular) or, sometimes, s lines. A Zeeman triplet agreeing thus in all respects, even in spacing, with the classical theory is called a *normal triplet*.

Concerning the *relative intensity* of the lines, wave mechanics makes the following predictions. Suppose all of the initial quantum states are excited equally, i.e., equal numbers of atoms are thrown into each of these states per second by the process of excitation. Under these conditions, the two σ lines will each be half as bright as the central π line, as seen transversely to the magnetic field; whereas, the σ lines will be twice as bright when viewed along the field as they are when viewed transversely. This latter result is easily seen to be in agreement with the classical theory of the Zeeman effect; each circular motion is equivalent to two linear vibrations at right angles to each other, but, in the transverse direction, radiation is received from only one of these, the one that lies in the line of sight being invisible.

The quantity $eH/(4\pi mc)$, representing the separation in frequency of each outer line from the central line in a normal triplet due to the field H , is often called a *Lorentz unit*, symbol L . Measured in waves per cm., its value is

$$\tilde{L} = \frac{e}{4\pi mc^2} H = 4.67 \times 10^{-5} H \text{ cm.}^{-1} \quad (219)$$

Thus, the normal separation is $4.67 \times 10^{-5} \text{ cm.}^{-1}$ per oersted. We note that L does not contain Planck's constant h . It is because of

this circumstance that classical theory gave correct results for certain cases of the Zeeman effect.

Up to this point, we have completely neglected the spin-orbit terms in the wave equation. Their effect, when the magnetic field is very intense, is only to shift the levels somewhat, in such a way as to separate slightly in energy those quantum states which have the same value of $\lambda + 2\mu$ but different values of $m = \lambda + \mu$. For example, states with $\lambda = 1$ and $\mu = -\frac{1}{2}$ and with $\lambda = -1$ and $\mu = \frac{1}{2}$ both have $\lambda + 2\mu = 0$ and so fall together if the spin-orbit effect is omitted entirely; but $m = \frac{1}{2}$ for the first and $m = -\frac{1}{2}$ for the second, so the energy levels for these two states are in reality slightly separated. The number of lines in the Zeeman pattern is thereby increased; but they fall into three groups in the approximate positions of the normal triplet.

Experimental confirmation of these conclusions from the theory is not extensive for 1-electron atoms, because very intense fields are required. However, the first three Balmer lines of hydrogen were observed as approximate normal triplets by Paschen and Back in a field of 26,900 oersteds.¹ The fine structure due to the spin-orbit effect was not resolved, but it was presumably responsible for the broadness of the observed lines.

On an atom containing more than 1 electron, the effect of a huge magnetic field is to split each energy level W_0 for zero field according to the more general equation

$$W = W_0 + \frac{eh}{4\pi mc}[\Sigma(\lambda + 2\mu)]H, \quad (220)$$

which constitutes an obvious generalization of Eq. (216); here the summation extends over all electrons in the atom. The normal triplet occurs as with 1 electron, since only 1 electron can jump at a time. The selection rules now include the rule that

$$\Delta M = 0 \text{ or } \pm 1, \quad (221)$$

where $M = \Sigma(\lambda + \mu)$ and denotes the total component of angular momentum in the direction of the field.

To obtain a good approximation to this type of Zeeman pattern with an atom containing more than 1 electron, the magnetic field must be so strong that it swamps all other influences except those for which allowance is made in the central field. The necessary strength runs to millions or even hundreds of millions of oersteds, so that Zeeman patterns of this type cannot be observed at present.

¹ PASCHEN and BACK, *Ann. d. Physik*, vol. 39, p. 897 (1912).

158. Zeeman Effect in a Weak Field.—The Zeeman effect is comparatively simple when the magnetic field is either *tremendously large* or *extremely small*. It will be convenient to take up next the case of a “weak” field, *i.e.*, a field weak enough so that the Zeeman splitting is small as compared even with the separations in the ordinary fine structure. In such a field, J is a good quantum number and the effect of the field is merely to separate the states of different M that compose each J level.

According to Eq. (212), the effect of a magnetic field upon the energy is determined by the average values of the electronic orbital and spin momenta in the direction of the field. The most important results obtained from wave mechanics in regard to these average

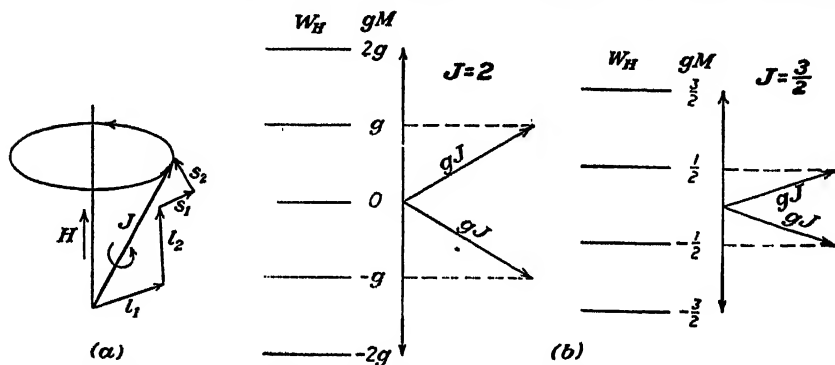


FIG. 114.—Diagram illustrating (a) vector addition in a magnetic field and (b) the separation of the magnetic levels in a weak field.

values can readily be inferred from the vector model that we have so often mentioned (Sec. 140). To do this, we think of the various electronic angular momenta as added together vectorially into a resultant angular momentum G_J . A diagram can be drawn, as in Fig. 114(a), showing the various L - and S -vectors added into a resultant J -vector; and we can then imagine the figure to whirl at arbitrary speed about the direction of the J -vector. Thereby we visualize the fact that, according to wave mechanics, the components of the electronic momenta perpendicular to G_J average to zero; the average values of these momenta thus lie along G_J . If components along G_J are then projected, in their turn, upon the direction of the magnetic field, we obtain the average components of the electronic momenta in the direction of the field. For the component of any electronic momentum perpendicular to G_J will contribute nothing to the average in the direction of the field, since it takes on equally often all directions about G_J . In this way can be obtained the average total orbital

momentum in the direction of the field, which we have denoted by \tilde{G}_{IH} , and the corresponding component of the spin momentum, \tilde{G}_{sH} .

The whole figure can then be supposed to whirl about the direction of the field at a fixed angle θ . (According to classical theory a rotation of this sort actually occurs, owing to the Larmor precession.) The projection of any vector parallel to G_J is then proportional to $\cos \theta$; hence so are \tilde{G}_{IH} and \tilde{G}_{sH} . But $\cos \theta$, in turn, is proportional to the component of G_J in the direction of H and, therefore, to the quantum number M (Sec. 140 and Fig. 94). Thus we can write

$$\tilde{G}_{IH} = g_1 \frac{Mh}{2\pi}, \quad \tilde{G}_{sH} = g_2 \frac{Mh}{2\pi},$$

and

$$\sum \tilde{G}_{IH} + 2 \sum \tilde{G}_{sH} = g \frac{Mh}{2\pi}, \quad (222)$$

where g_1 and g_2 are constants of proportionality (independent of M), and $g = \Sigma g_1 + 2 \Sigma g_2$, Σ indicating summation over all electrons in the atom.

Substituting the value given by (222) for the quantity in parentheses in Eq. (211a), we have for the magnetic moment of the atom in the direction of the field, when it is in one of the M states,

$$\mu = - \frac{eh}{4\pi mc} g M. \quad (223)$$

Making the same substitution in the expression given in Eq. (212) for the energy, and adding W_{0J} , representing the energy of the J level in the absence of a magnetic field, we have for the energy

$$W_{JM} = W_{0J} + \frac{eh}{4\pi mc} g M H. \quad (224)$$

(Here, as usual, e stands for the numerical electronic charge.)

The last equation shows that the "magnetic" states composing a given J level and corresponding to different values of M become separated and equally spaced in a weak magnetic field, but the spacing differs from that corresponding to classical theory, as represented by Eq. (220), by a factor g . The value of g may vary from one J level to another. It is often called the *Landé splitting factor* or *g-factor*. The splitting can be visualized by means of an obvious modification of the vector model, as in Fig. 114(b), vectors being drawn of length gJ instead of J .

A more convenient equation for spectroscopic purposes is obtained if we divide all energies by hc so as to express them in equivalent

wave numbers. At the same time, let us introduce the Lorentz unit defined above,

$$\tilde{L} = \frac{eH}{4\pi mc^2} = 4.67 \times 10^{-5} H \text{ cm.}^{-1}, \quad (225)$$

representing the normal separation of either σ line from the π line. Then (224) becomes

$$\tilde{W}_{JM} = \tilde{W}_{0J} + gM\tilde{L}. \quad (226)$$

For the values of \tilde{W}_{JM} for two different J levels, we may then write

$$\tilde{W}_{0J'} + g'M'\tilde{L}, \quad \tilde{W}_{0J''} + g''M''\tilde{L}.$$

Subtracting these two expressions, we have then for the frequencies of the Zeeman lines between the two J levels in question, all of which would coalesce into one line if the magnetic field were absent:

$$\bar{\nu} = \bar{\nu}_0 + (g'M' - g''M'')\tilde{L}, \quad (227)$$

where $\bar{\nu}_0 = \tilde{W}_{0J'} - \tilde{W}_{0J''}$ and represents the frequency when $H = 0$. If J' , J'' are the corresponding values of J for the two levels, in Eq. (227) M' takes on integrally spaced values from $-J'$ to J' , and M'' from $-J''$ to J'' , subject only to the selection rule

$$\Delta M = M'' - M' = 0 \text{ or } \pm 1.$$

As in Sec. 157, transitions in which $\Delta M = 0$ give rise to π lines; those in which $\Delta M = \pm 1$, to σ lines. The arrangement of the components into which a spectral line is split by a magnetic field is called the "Zeeman pattern" of the line.

All observed Zeeman patterns in weak fields are found to agree with Eq. (227) if suitable values of g' and g'' are inserted in it. If $g' \neq g''$, the observed Zeeman pattern leads to a knowledge not only of g' and of g'' but also of the values of J for both levels, since the number of magnetic states for given J is $2J + 1$. (The student may be interested to invent a method for inferring values of J and g from such observations.) Thus, the Zeeman effect is of great utility in the analysis of spectra. A mathematical description of the anomalous Zeeman effect, in terms of an equation equivalent to (227), was worked out empirically by Landé before the discovery of wave mechanics.

To calculate a *theoretical* value of g requires, in general, a knowledge of the wave function. If *LS coupling occurs*, however, wave mechanics furnishes a simple general formula for g ; and it happens that this formula can be inferred by resorting again to the vector diagram.

For this purpose, we shall suppose the diagram to represent the angular momenta in full detail (*cf.* Sec. 140). The orbital momenta of the individual electrons are first added into a resultant orbital angular momentum \mathbf{G}_L , about which we suppose them to whirl (Fig. 115); their components perpendicular to \mathbf{G}_L will thus average to zero, while their components along \mathbf{G}_L will just add up to the magnitude of \mathbf{G}_L itself, which we may denote by G_L . The average component of $\Sigma \mathbf{G}_I$ along \mathbf{G}_J will thus be $G_L \cos \theta_1$ (see figure), by the argument used above. Imagining \mathbf{G}_J then to whirl about the direction of H , we have, for the average component of $\Sigma \mathbf{G}_I$ in the direction of H ,

$$\Sigma \bar{G}_{IH} = G_L \cos \theta_1 \cos \theta,$$

θ_1 and θ being the fixed angles between \mathbf{G}_J and \mathbf{G}_L and between \mathbf{G}_J and H , respectively (Fig. 115). Similarly,

$$\Sigma \bar{G}_{SH} = G_S \cos \theta_2 \cos \theta,$$

θ_2 being the angle between \mathbf{G}_J and \mathbf{G}_S . But, by elementary geometry,

$$\cos \theta_1 = \frac{1}{2G_J G_L} (G_J^2 + G_L^2 - G_S^2), \quad \cos \theta_2 = \frac{1}{2G_J G_S} (G_J^2 + G_S^2 - G_L^2);$$

and

$$\cos \theta = \frac{Mh}{2\pi} \frac{1}{G_J},$$

since $Mh/2\pi$ is the component of \mathbf{G}_J parallel to H . Hence,

$$\sum \bar{G}_{IH} = \frac{Mh}{4\pi G_J^2} (G_J^2 + G_L^2 - G_S^2), \quad \sum \bar{G}_{SH} = \frac{Mh}{4\pi G_J^2} (G_J^2 + G_S^2 - G_L^2).$$

We now substitute here

$$G_J^2 = \frac{J(J+1)h^2}{4\pi^2}, \quad G_L^2 = \frac{L(L+1)h^2}{4\pi^2}, \quad G_S^2 = \frac{S(S+1)h^2}{4\pi^2}.$$

(*Cf.* Secs. 140, 141.) Thus, we find

$$\begin{aligned} \sum \bar{G}_{IH} + 2 \sum \bar{G}_{SH} &= \frac{Mh}{4\pi J(J+1)} [3J(J+1) + S(S+1) - L(L+1)] \\ &= \frac{Mh}{2\pi} \left[1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \right]. \end{aligned}$$

Comparing this with Eq. (222), we see that, in case LS coupling is

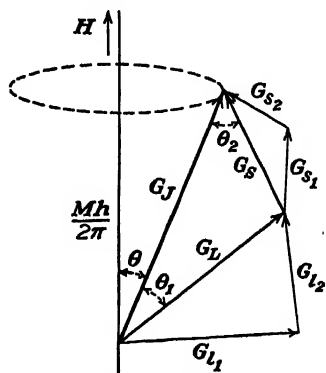


FIG. 115.

valid,

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \quad (228)$$

Theoretical formulas for g in case jj coupling holds can also be worked out.¹ To calculate g from theory in other cases is more difficult.

The values of g inferred from observation agree, in general, fairly well with the theoretical formulas whenever the coupling is shown by other considerations to approximate to the type assumed.

159. Zeeman Patterns of LS Multiplets in a Weak Field.—Equation (227) shows that the splitting of a spectral line in a weak field will depend upon the values of g for both of the levels involved. A wide variety of patterns is possible. Since the various component lines of an LS spectral multiplet arise from levels varying in the value of J , these levels will also have different g values, in general, and the lines of the multiplet will, therefore, exhibit different Zeeman patterns. *Corresponding lines* in the spectral terms of the *same series*, however, should exhibit the same type of pattern; for such lines have a common final level, and their initial levels have the same J , S , and L , and hence the same value of g . This conclusion from the theory is in agreement with observation; it is known as "Preston's rule," having been discovered empirically by Preston in 1898. Preston's rule is sometimes of use in deciding what lines belong together in a series.

The following particular cases deserve comment.

(a) $g' = g'' = 0$. If $g = 0$ both for the initial and for the final state, we see from Eq. (227) that the line is *not split at all* in a weak field. Such cases can arise in several ways. *Every level with $J = 0$ can be supposed to have $g = 0$.* Actually, in Eq. (228) the fraction is indeterminate; but if $J = 0$, there is only one magnetic level, hence no splitting of the J level; and the single level with $M = 0$, by Eq. (224), is not even displaced in a weak field. It is customary to write $g = \infty$ when $J = 0$, but it seems simpler just to write $g = 0$. A jump between two levels with $J = 0$ is excluded, to be sure, by the usual selection rule; but it can also happen that $g = 0$ when $J \neq 0$, e.g., for a $^4D_{1/2}$ level ($J = \frac{1}{2}$, $S = \frac{3}{2}$, $L = 2$) or a 5F_1 ($J = 1$, $S = 2$, $L = 3$). A line of the type $^5F_1 \rightarrow ^5D_0$ should, therefore, show no Zeeman effect at all. This is confirmed by experiment (e.g., the line $\lambda = 5,713$ Å in the arc spectrum of titanium).

(b) If $g' = g'' = g \neq 0$, or if $J = 0$ for one level only and $g \neq 0$ for the other, the Zeeman pattern is a *triplet*. For, if $J = 0$ for the

¹ Cf. WHITE, *op. cit.*

second level, in Eq. (227), $M'' = 0$ and

$$\tilde{\nu} = \tilde{\nu}_0 + g'M'\tilde{L}.$$

Since $M' - M'' = 0$ or ± 1 , $\Delta M' = 0$ or ± 1 here, and three lines are obtained. Of, if both levels have the same $g \neq 0$,

$$\tilde{\nu} = \tilde{\nu}_0 + g(M' - M'')\tilde{L};$$

the selection rule for M gives again three lines, although in this case they may be obtained several times over, as in the huge-field effect. The outer or σ lines are displaced a distance $g\tilde{L}$ from the central π line.

If also $g = 1$, the triplet is a *normal* one (Sec. 157), spaced as in classical theory. The value $g = 1$ always occurs when $S = 0$, so that $J = L$. Thus all lines in a *singlet spectrum* exhibit the *normal Zeeman* effect in a weak field. An example is the singlet spectrum of mercury (Sec. 151). In singlet levels the electron spins are opposed and so produce no magnetic effect, whereas the orbital effect by itself produces the classical pattern, as we found to be the case in a huge field (Sec. 157). Zeeman had the good fortune to succeed first in resolving the components in singlet lines, *e.g.*, in the spectra of zinc, cadmium, or tin.

A few of the theoretical values of g for LS levels are shown in Table VIII.

TABLE VIII.—THE LANDÉ SPLITTING FACTOR g FOR LS COUPLING

S	$J \backslash L =$	$S \quad 0$	$P \quad 1$	$D \quad 2$	$F \quad 3$	$G \quad 4$
0	L	0	1	1	1	1
$\frac{1}{2}$	$L - \frac{1}{2}$..	$\frac{2}{3}$	$\frac{4}{5}$	$\frac{6}{7}$	$\frac{8}{9}$
	$L + \frac{1}{2}$	2	$\frac{4}{3}$	$\frac{6}{5}$	$\frac{8}{7}$	$\frac{10}{9}$
1	$L - 1$..	0	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{5}{6}$
	L	..	$\frac{3}{2}$	$\frac{5}{6}$	$\frac{7}{4}$	$\frac{9}{2}$
	$L + 1$	2	$\frac{5}{2}$	$\frac{7}{3}$	$\frac{9}{4}$	$\frac{11}{2}$

Some typical spectral patterns, as predicted by the theory, are shown in Fig. 116. Examples of the anomalous Zeeman effect are shown in all cases but one. The lengths of the vertical lines as drawn indicate the relative theoretical intensities of the spectral lines as seen in a direction perpendicular to the field, which are usually in agreement with experiment; π lines are drawn above, σ lines below the horizontal axis. Three dots indicate the normal triplet in the same field. In each case the two J levels involved in the transition are indicated.

The student will find it instructive to check the spacing from the formulas in a number of cases. Other diagrams of the same sort are shown in H. E. White's book.¹

To illustrate the use of the theoretical equations, consider the familiar *D* lines of sodium, which arise from transitions of the type

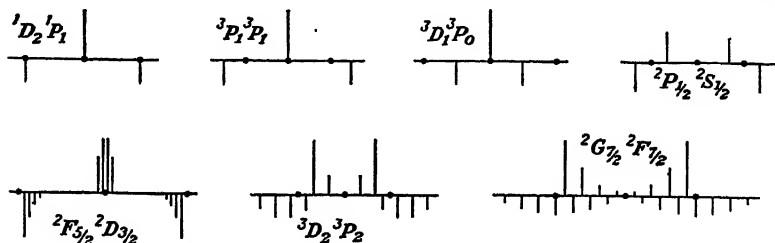


FIG. 116.—Some Zeeman patterns for *LS* coupling in a weak field. Heights of lines indicate roughly the relative intensities of the lines. π lines are drawn above, σ lines below. The three dots indicate the position of the normal triplet in the same field.

${}^2P_{3/2} \rightarrow {}^2S_{1/2}$ and ${}^2P_{1/2} \rightarrow {}^2S_{1/2}$. For the ${}^2P_{3/2}$ state, we insert in Eq. (228) $J = 3/2$, $S = 1/2$, $L = 1$, obtaining

$$g = 1 + \frac{\frac{3}{2}(\frac{3}{2} + 1) + \frac{1}{2}(\frac{1}{2} + 1) - 1(1 + 1)}{2 \times \frac{3}{2}(\frac{3}{2} + 1)} = \frac{4}{3}.$$

Since M may take on any one of the four values $3/2, 1/2, -1/2, -3/2$, gM in Eq. (224) or (226) may have any one of the four values

$$\frac{4}{3}, \frac{2}{3}, -\frac{2}{3}, -\frac{4}{3}.$$

The level ${}^2P_{3/2}$ is thus split into four magnetic levels removed from the position of the level in zero field by gM Lorentz units or $gM\tilde{L}$ cm.⁻¹. Similarly, the state ${}^2P_{1/2}$ is split into two levels for which

$$g = 1 + \frac{\frac{1}{2}(\frac{1}{2} + 1) + \frac{1}{2}(\frac{1}{2} + 1) - 1(1 + 1)}{2 \times \frac{1}{2}(\frac{1}{2} + 1)} = \frac{2}{3}, \quad gM = \pm \frac{1}{3};$$

and the ${}^2S_{1/2}$ level is split into two with

$$g = 1 + \frac{\frac{1}{2}(\frac{1}{2} + 1) + \frac{1}{2}(\frac{1}{2} + 1)}{2 \times \frac{1}{2}(\frac{1}{2} + 1)} = 2, \quad gM = \pm 1.$$

Figure 117 shows these levels, and also the transitions between them that are allowed by the selection rule $\Delta M = 0$ or ± 1 . The resulting line patterns are shown at the bottom of the figure. The expected displacement of each component line from the position for zero field is calculated, in Lorentz units, by subtracting the two gM values; e.g.,

¹ *Op. cit.*, pp. 161, 222.

for the left-hand σ component of D_2 , we have $-\frac{2}{3} - 1 = -\frac{5}{3}$ Lorentz units, which is shown below the line in an obvious notation. Further explanation is added below the figure.

We note that D_1 or ${}^2P_{3/2} \rightarrow {}^2S_{1/2}$ should be resolved into four components, two π lines ($\Delta m = 0$) polarized parallel to the field and two σ lines ($\Delta m = \pm 1$) polarized perpendicularly. D_2 or ${}^2P_{3/2} \rightarrow {}^2S_{3/2}$ should be resolved into six components of which the outer four are σ lines. A glance at the photograph in Fig. 118 shows that the theoretical predictions are confirmed by experiment, as to the patterns

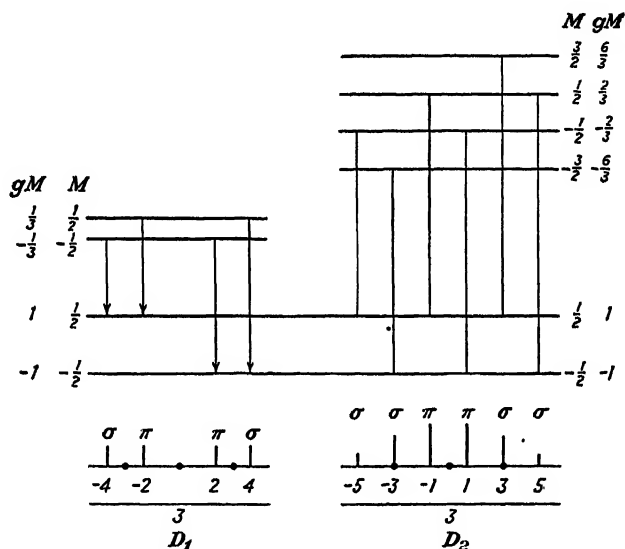


FIG. 117.—Weak-field Zeeman effect for the D lines of sodium. The same scale is used for all of the Zeeman separations, also for the theoretical relative intensities of the component lines as shown by the relative heights of the lines as drawn. The position of the normal triplet in the same field is shown by three dots.

themselves, and at least roughly also as to intensities. Examination of the lines through a Nicol prism shows that the polarizations also are correctly predicted.

160. The Paschen-Back Effect.—In the discussion in the last two sections the emitting atom has been assumed to be in a “weak” magnetic field. The term “weak” is, of course, relative, since fields in excess of 10,000 oersteds are usually necessary to produce observable Zeeman patterns. A magnetic field is said to be “weak” (or strong) according as the separations of the Zeeman components of each of the several lines of a multiplet group are small (or large), compared with the separation of the lines of the multiplet from one another. Figure 119, for example, shows diagrammatically to scale the D lines of sodium

with their respective Zeeman patterns in a field of 30,000 oersteds. The Zeeman separations are seen to be small compared with the separation between D_1 and D_2 . For the D lines of sodium, therefore, a field of 30,000 oersteds is a "weak" field.

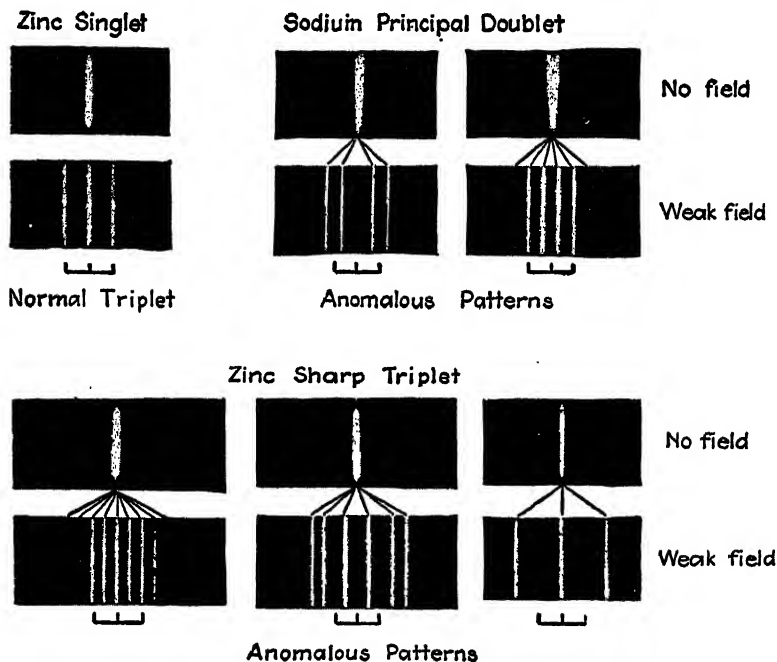


FIG. 118.—Photographs of some Zeeman patterns in a weak magnetic field, viewed perpendicularly to the field. The brackets show in each case the position of a normal triplet in the same field. (From H. E. White, "Introduction to Atomic Spectra.")

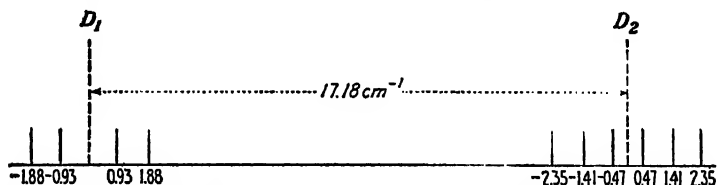


FIG. 119.—Relative separations of the D lines of Na(11) and their Zeeman components in a field of 30,000 oersteds.

Consider, however, the Zeeman effect in lithium. The separation of the two components of the first line of the principal series¹ is of the order of 0.3 cm^{-1} . A field of 30,000 oersteds would be expected to produce Zeeman separations several times larger than this doublet

¹ See Table VII, p. 365.

separation [according to (225), if $H = 30,000$, $\tilde{L} = 1.4 \text{ cm.}^{-1}$]. For lithium, therefore, a field of 30,000 oersteds is a "strong" field.

It was found by Paschen and Back¹ that, whatever the Zeeman patterns of a given LS multiplet of lines may be in a weak field, the Zeeman pattern of the group as a whole in a strong field always approximated a normal triplet. The transition from weak to strong field is an interesting phenomenon. As the field is increased to such a point that the magnetic levels into which one J level is split begin to approach those belonging to another J level, departures from the weak-field arrangement begin to occur. As the field increases further, magnetic sublevels belonging to different J levels may cross one other, provided they do not have the same value of M . Eventually, when the field has become strong enough, the magnetic levels are observed to group themselves more or less closely into a set of equally spaced composite levels, and the spectral pattern approximates a normal triplet; i.e., we have then the case of a huge field as described in Sec. 157.

The transition from one type of Zeeman effect to the other is accompanied by changes in the angular momenta. As long as the field is weak, the resultant orbital momentum \mathbf{G}_L and the resultant spin momentum \mathbf{G}_S remain added approximately into a resultant momentum \mathbf{G}_J of fixed magnitude, characterized by a certain value of the quantum number J . As the field grows stronger, however, this coupling is progressively broken down, and eventually, in a strong field, it disappears. Then Λ and Σ become good quantum numbers, along with L and S ; the resultant orbital momentum \mathbf{G}_L by itself has a fixed component in the direction of the field of magnitude $\Lambda h/2\pi$, and \mathbf{G}_S has similarly a component $\Sigma h/2\pi$. The total angular momentum in the direction of the field is thus $Mh/2\pi$, where $M = \Lambda + \Sigma$. Under these conditions the quantum number J ceases to have any significance at all. The theoretical treatment of the energy then parallels closely the theory for 1 electron in a huge field, as described in Sec. 157. The energy can be written, approximately, in analogy with Eq. (216),

$$W = W_0 + \frac{eh}{4\pi mc} (\Lambda + 2\Sigma) H,$$

W_0 being a constant for a given LS term. The selection rules are, besides those for L and S ,

$$\Delta\Lambda = 0 \text{ or } \pm 1, \quad \Delta M = 0 \text{ or } \pm 1, \quad \Delta\Sigma = 0.$$

The normal spectral triplet is produced in essentially the same way as for 1 electron.

¹ *Ann. d. Physik*, vol. 39, p. 897 (1912); vol. 40, p. 960 (1913).

In this discussion, however, we have ignored the spin-orbit and additional electrostatic effects. These effects actually separate slightly those levels which have the same value of $\Lambda + 2\Sigma$ but different values of M and thereby introduce a sort of fine structure into the Zeeman pattern.

A corresponding theory is easily developed for the case of jj coupling. In this case, however, or in any case other than that of LS coupling, there is an intermediate stage in the progressive change of the Zeeman pattern.

All of these changes of pattern as the strength of the magnetic field increases are known collectively as the *Paschen-Back effect*. The effect is said to be *incomplete* when the magnetic field is strong enough to make the Zeeman splitting for the energy levels larger than the mutual separation of the J levels composing a given LS or jj term, or a similar group in other cases, but yet small as compared with the separations between the terms themselves. Such a field is called merely "strong." If the field could be increased still further, until the magnetic levels of two different *terms*, such as a 1P and a 1D , approached each other, the pattern should change again, approaching the normal triplet as the field becomes so large that its effect predominates over everything else except the central field. This stage, described in Sec. 157 as the effect in a "huge field," is known as the *complete* Paschen-Back effect. Unfortunately, it can scarcely be realized except in 1-electron atoms, in which the first two stages of the Zeeman effect are practically unobservable.

161. The Stark Effect.—An effect of an electric field, somewhat analogous to the Zeeman effect, was discovered by Stark in 1913 and is known as the "Stark effect." It is observed in the well-known "canal rays," when these are moving in an electric field of several thousand volts per centimeter. An arrangement for producing the Stark effect is shown diagrammatically in Fig. 120. The anode and cathode in a glass tube are A and C , respectively, in which the gas is maintained at such a pressure that the Crookes dark space in front of C is several centimeters long. The cathode C is perforated with small holes through which pass, in cylindrical bundles, luminous streams of atoms of the gas that have acquired a positive charge immediately in front of C . These streams of atoms are the canal rays. A third terminal S is placed immediately behind C , at a distance of a few millimeters, and an electric field of several thousand (20,000 and up) volts per centimeter is maintained between S and C . For the "transverse" Stark effect, light from the canal rays enters the spectroscope in the direction of the arrow.

Stark observed that the spectral lines emitted by the canal rays when no field exists between S and C are split up, when the field is applied, into numerous components which are polarized, some of them parallel to the field (p - or π -) components, others perpendicular (s - or σ -components). The splitting is usually proportional, at first, to the square of the electric field strength; in an intense field, however, the pattern becomes quite different, in analogy with the Paschen-Back effect in a magnetic field.

Classical theory could not account for the Stark effect, but a good account of it was developed by Epstein, using an extension of Bohr's theory; and it is completely explained by wave mechanics. The

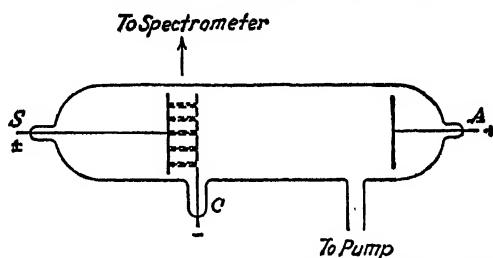


FIG. 120.—Arrangement for producing the Stark effect.

treatment of the Stark effect constituted, in fact, one of the most striking successes of quantum theory. In spectroscopy, however, this effect is of minor interest because it does not, like the Zeeman effect, serve to illuminate other parts of the subject. Examples of Stark patterns are given in White's book.¹

162. The Stern-Gerlach Experiment.—The space quantization of atomic angular momenta, upon which we have based the discussion of the Zeeman effect, constitutes a characteristic feature both of the older quantum theory and of wave mechanics, in contrast with classical theory. Its occurrence is not definitely confirmed by the existence of the Zeeman effect, however, in view of the partial success of classical theory in accounting for these phenomena. An experiment which directly reveals the space quantization itself was proposed by O. Stern in 1921 and was carried out by him in collaboration with Gerlach.²

In elementary magnetic theory, it is shown that a magnet tends to move so as to increase the magnetic flux through it in the direction of its magnetic axis. In a uniform field, the only result is that the magnet experiences a torque tending to line it up with the field. In a

¹ *Op. cit.*; cf. also HERZBERG, *op. cit.*

² STERN, *Zeits. f. Physik*, vol. 7, p. 249 (1921); GERLACH and STERN, *Zeits. f. Physik*, vol. 8, p. 110; vol. 9, p. 349 (1922); *Ann. d. Physik*, vol. 74, p. 673 (1924).

nonuniform field, however, the magnet experiences a *translatory force* as well.

Suppose now a slender beam of atoms having magnetic moments travels in the direction of the x -axis across a magnetic field whose lines are approximately parallel to the y -axis but whose magnitude increases rapidly in the direction of the z -axis. After passing through the field, let these atoms be collected on a suitable target. Then atoms which have a component of their magnetic moment in the direction of the field will be deflected sideways; if the component of the moment has the same direction as the field, they will be deflected toward $+z$; if it has the opposite direction, toward $-z$.

So far, the predictions of classical and quantum theory coincide. According to *classical* theory, however, each atom will enter the field

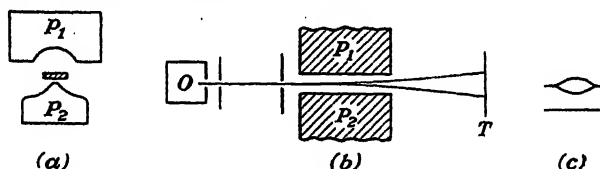


FIG. 121.—Illustrative of the Stern-Gerlach experiment. (a) and (c) represent views seen in the direction of the beam; (c) shows the trace on the target (exaggerated), with and without the magnetic field.

with its magnetic axis inclined at some angle θ to the field, and the axis will then execute a Larmor precession about the field at the fixed angle θ . Since all values of θ will occur among the atoms, their deflections will be distributed in continuous fashion, and the atoms, instead of forming a small spot on the target, will be drawn out into a continuous band.

According to *quantum* theory, on the other hand, each atom will enter the field in a certain quantum state, defined with the direction of the field as an axis. Its magnetic moment in the direction of the field will be gM Bohr magnetons (if the field is not too strong), where g is the Landé splitting factor and M has one of several integral values [cf. Eq. (223)]. The beam will be broken up, therefore, into separate beams and will form on the target a series of distinct spots, one for each possible value of M .

The arrangement used in the experiment of Stern and Gerlach is shown diagrammatically in Fig. 121. The nonhomogeneous field was produced between pole pieces of which one had a sharp edge, so that near it the field was much stronger than elsewhere [cf. Fig. 121(a)]. A strap-shaped beam of silver atoms was formed by evaporating silver in a heated oven O and allowing atoms from the vapor to stream out through collimating slits; the beam [shown by the small rectangle in

Fig. 121(a)] traveled closely past the sharp edge of pole piece P_2 and was condensed on a plate at T . With no field, the beam formed a narrow line on the plate [Fig. 121(c), lower line]. When the magnetizing current was turned on, the line was not widened continuously but was divided into two lines, as shown in the upper part of Fig. 121(c), except at the ends, which were due to atoms passing at some distance from the sharp edge.

Space quantization of the silver atoms was thus clearly revealed. From careful measurements of the separation of the two lines and of the gradient of the magnetic field strength, it was calculated that each silver atom had a magnetic moment in the direction of the field of 1 Bohr magneton, with an error of 10 percent at most.

These results, and others obtained subsequently, are in complete agreement with the predictions of wave mechanics. The silver atom is normally in a $^2S_{1/2}$ state, for which $g = 2$; thus, half of the atoms should have $M = 1/2$ and a moment in the direction of the field of $2 \times 1/2 = 1$ Bohr magneton; whereas, for the other half, $M = -1/2$ and the moment is -1 . Similar results were obtained subsequently, with greater precision, by Taylor¹ and by Leu² for sodium and potassium, which are likewise in $^2S_{1/2}$ states. Atoms of zinc and cadmium, in Leu's experiments, were unaffected by the magnetic field. Atoms of thallium gave a double line corresponding to a moment of $1/3$ magneton. Zinc and cadmium are normally in 1S_0 states, which have $M = 0$ and no magnetic moment. For thallium the normal state is inferred from spectroscopic evidence to be $^2P_{1/2}$; thus $M = 1/2$ or $-1/2$ again, but $g = 2/3$ instead of 1 [by Eq. (228), in which $J = 1/2$, $S = 1/2$, $L = 1$].

The method of magnetic deflections has been extended very ingeniously by Rabi for the measurement of the magnetic moments of *nuclei*. The student will find it worth while to look up his papers on this subject in the *Physical Review*, vol. 49 onward.

163. Isotope Structure and Hyperfine Structure.—The ordinary fine structure due to spin-orbit interaction does not exhaust the possibilities of fine details in spectral lines. Even before 1900, Michelson and others had shown, by means of the interferometer, that many spectral lines possess a further structure much finer still. This came to be known as hyperfine structure.

After the discovery that many chemical elements consist of several isotopes (Sec. 206), it was believed for a time that each component of a hyperfine pattern of lines was emitted by a different isotope. But

¹ TAYLOR, *Phys. Rev.*, vol. 28, p. 276 (1926).

² LEU, *Zeits. f. Physik*, vol. 41, p. 551 (1927).

later, hyperfine structure was discovered in the spectra of some elements which, as shown by the mass spectrograph, consist of only 1 isotope. An example is bismuth; the line $\lambda = 3,596 \text{ \AA}$ contains six hyperfine components spread over a range of 0.3 \AA or 2.3 cm^{-1} . It was suggested by Pauli in 1924 that this effect might be due to an occurrence in the nucleus of angular momentum and an associated magnetic moment. It appears that both causes are operative. To distinguish between them, a tendency has arisen to restrict the term *hyperfine structure* to that which is associated with the angular momentum of the nucleus, the other type being called *isotope structure*.

A photograph illustrating *isotope structure* in the spectrum of tungsten is shown in Fig. 122. The spectrum was formed with a

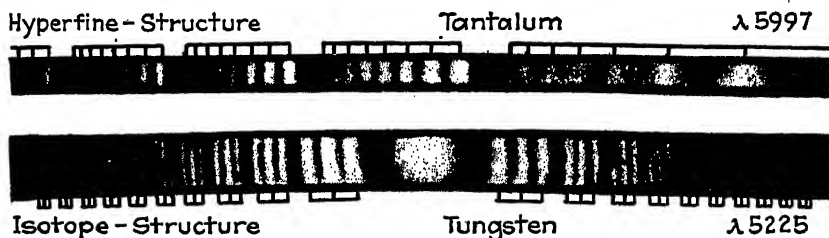


FIG. 122.—Photographs illustrating (a) hyperfine structure in a spectral line of tantalum and (b) isotope structure in a line of tungsten. (After Grace, More, MacMillan, and White; from H. E. White, "Introduction to Atomic Spectra.")

Fabry-Perot etalon, hence the same pattern appears repeated many times in different orders. Tungsten consists of 4 isotopes, with mass numbers 182, 183, 184, 186, the first two being somewhat less abundant than the last two. The three lines observed were ascribed to the 3 isotopes of even number, the line due to 183 being assumed to be masked by the others.

Isotope shifts of spectral lines are found, in general, to be proportional, for a given element, to the differences in atomic mass. As a rule, however, they do not seem to be purely the result of nuclear motion, as is the difference in the positions of those lines in the spectra of hydrogen and of ionized helium which would otherwise fall together (Sec. 106). The shift associated with an increase in mass from one isotope to another is even sometimes in opposite directions in different elements. The cause of most isotope shifts is thought to lie, perhaps, in some departure from the Coulomb type of field very close to the nucleus, the departure being different for different elements or even for different isotopes of the same element.

An interesting example in which the isotope shift probably is a simple mass effect is afforded by the spectrum of hydrogen, in which it

was used as a guide by Urey and his collaborators in the discovery of heavy hydrogen or deuterium.¹ The Rydberg constants for the two kinds of hydrogen atoms can be found by substituting $M = M_1$ and $M = 2M_1$ successively in Eq. (146a) in Sec. 106, which gives

$$R_1 = \frac{R_\infty M_1}{M_1 + m}, \quad R_2 = \frac{2R_\infty M_1}{2M_1 + m},$$

M_1 being the mass of an ordinary hydrogen atom and m the mass of an electron. The difference in wave length for a given line of wave length λ will then be, nearly enough,

$$\Delta\lambda = - \frac{\lambda(R_2 - R_1)}{R_\infty} = - \frac{\lambda m}{2M_1}.$$

Since $M_1/m = 1,837$, this means that, for example, the $H\beta$ line for deuterium should lie $4,861/3,674 = 1.32$ Å on the violet side of that for ordinary hydrogen atoms. This line was observed to be very faintly visible, in the expected position, in the spectrum from a sample of common hydrogen, and it increased in strength as the hydrogen was subjected to operations which should increase the relative concentration of deuterium.

The theoretical treatment of *true hyperfine structure*, on the other hand, resembles closely the treatment of *L*-*S* fine structure due to the spin-orbit effect, as described in Secs. 144 and 145. The total angular momentum of the *nucleus* is assumed to be fixed, so that it can be represented by a quantum number I . The value of I may be 0 or integral or half-integral, and it may be different for different isotopes of the same chemical element. The angular momentum in question is presumably the resultant of spin momenta of the protons and neutrons of which the nucleus is composed; hence, it is commonly said to be due to nuclear "spin." I remains constant for a given atom because in ordinary physical processes the nucleus remains always in its lowest energy level. This level is degenerate, however, if $I > 0$; there is a second quantum number M_I , restricted to values such that $|M_I| \leq I$. Thus, I and M_I in nuclear theory correspond to J and M in atomic theory.

If the action of the nucleus on the electrons were purely that of a central field, the only effect of the nuclear angular momentum would be a $(2I + 1)$ fold degeneracy of all quantum states, with the selection rule $\Delta M_I = 0$. But there may exist also an associated nuclear magnetic moment, as in atoms. Equation (216) in Sec. 157 and its physical interpretation suggest that a single particle with a mass M equal to

¹ UREY, BRICKWEDDE, and MURPHY, *Phys. Rev.*, vol. 40, p. 1 (1932).

that of a proton, if, like an electron, it has spin $\frac{1}{2}$, might have an effective magnetic moment of magnitude

$$\frac{eh}{4\pi Mc} = \frac{1}{1,837} \text{ Bohr magneton.}$$

This is called a *nuclear magneton*. Observation indicates, however, that the magnetic moment of the proton or hydrogen nucleus is actually nearer to 3 nuclear magnetons.¹

If such a nuclear moment exists, wave mechanics indicates that the angular momentum of the nucleus is to be added to the resultant angular momentum of the electrons by a process of *IJ* coupling exactly analogous to the *LS* coupling of electronic orbital and spin momenta (Sec. 144). The result is then, in general, that each of the quantum states of the atom as dealt with hitherto is broken up into a group of quantum states characterized by two additional quantum numbers *F* and *M_F*, which have reference to the total angular momentum of the atom, that of the nucleus included.

The interaction between the nuclear magnetic moment and the electronic motions and spins then slightly separates those states which have different values of *F*. The number of hyperfine energy levels thus produced may be different for electronic states of different *J*. If $J \geq I$, their number is $2I + 1$; but, if $J < I$, it is $2J + 1$. For the spacing of the hyperfine levels the analog of Eq. (201) for *LS* multiplets is obtained; the separation of successive levels is proportional to the larger of the values of *F* for the two levels [cf. Eq. (202) in Sec. 146]. As selection rules, in addition to all others, we have

$$\Delta F = 0 \text{ or } \pm 1, \quad \Delta M_F = 0 \text{ or } \pm 1.$$

A *hyperfine spectral multiplet* results from transitions between the hyperfine levels composing two ordinary *J* levels. Usually it happens that the spacing of the hyperfine sublevels in one of the two *J* levels is much larger than that in the other, so that the former spacing stands out in the spectral multiplet as observed, the finer structure due to the other *J* level being frequently not resolved at all. The result is then an easily recognizable "flag" type of pattern, a good example of which is shown in Fig. 122.

As an illustration of the theory, consider the following doublet in the spectrum of thallium (Tl 81):

$$6s^2 6p \ ^2P_{1/2} \rightarrow 6s^2 7s \ ^2S_{1/2} \ (\lambda = 3,777 \text{ \AA}; \bar{\nu} = 26,478 \text{ cm.}^{-1}) \quad (a)$$

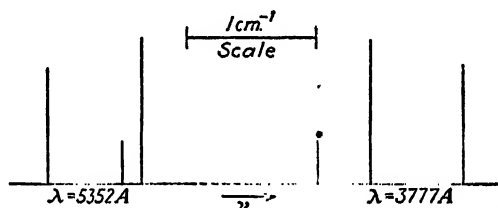
$$6s^2 6p \ ^2P_{3/2} \rightarrow 6s^2 7s \ ^2S_{1/2} \ (\lambda = 5,352 \text{ \AA}; \bar{\nu} = 18,684 \text{ cm.}^{-1}) \quad (b)$$

¹ KELLOGG, RABI, and ZACHARIAS, *Phys. Rev.*, vol. 50, p. 472 (1936).

TABLE IX.—HYPERFINE STRUCTURE OF THE COMPONENTS OF THE DOUBLET OF
 THALLIUM: $6s^2 6p \, ^2P_{1/2,3/2} \rightarrow 6s^2 7s \, ^2S_{1/2}$

λ	$\Delta\tilde{\nu}$, cm.^{-1}	Intensity	λ	$\Delta\tilde{\nu}$, cm.^{-1}	Intensity
3,776.888 Å		3	5,352.184 Å		8
	0.40			0.60	
3,776.830		10	5,352.014		3
	0.71			0.14	
3,776.729		8	5,351.974		10

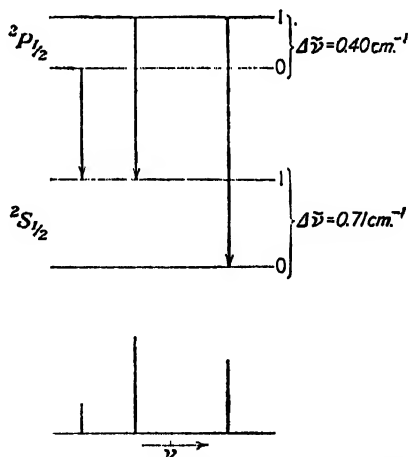
Under high resolution¹ each line of this doublet is seen to be made up of three components. Their wave lengths and estimated intensities


 FIG. 123.—Hyperfine structure of the thallium doublet $\lambda = \frac{5,352}{3,777}$ angstroms.

are given in Table IX; their relative positions on a frequency scale are shown in Fig. 123.

To produce three component lines, as observed, there must be two hyperfine sublevels in each J level, just as in the production of the $D \rightarrow P$ triplets in the alkali spectra there are two J levels in each LS term (Sec. 147). Hence, here $2 = 2I + 1$, and so $I = \frac{1}{2}$ for the thallium nucleus. Then for $J = \frac{1}{2}$, $F = I \pm \frac{1}{2}$ or $F = 1$ or 0 ; for $J = \frac{3}{2}$, $F = 2$ or 1 . The spacing of the hyperfine levels is easily determined from the observed separations in the spectrum. The level diagram is shown for one line of the doublet in Fig. 124.

It may be remarked that the same pattern would result if the doublet separations in Fig. 124 were


 FIG. 124.—Energy levels and transitions for the hyperfine structure of the thallium line $\lambda = 3,777$ angstroms.

¹ BACK, *Ann. d. Physik*, vol. 70, p. 367 (1923).

to be interchanged. According to Back and Wulff, the Zeeman pattern for the hyperfine components unambiguously assigns the energy-level differences.¹

In Table X are shown a few values of I and of the nuclear magnetic moment μ expressed in terms of nuclear magnetons. M is the mass number of the isotope in question (Sec. 206). Where no references are given, the data are taken from Condon and Shortley's book² and were calculated from the observed spectroscopic separations by means of perturbation theory.

TABLE X.—NUCLEAR SPIN NUMBERS I AND MAGNETIC MOMENTS μ

	Z	M	I	μ		Z	M	I	μ
H	1	1	$\frac{1}{2}$	2.78*	Cd	48	111, 113	$\frac{1}{2}$	-0.6
H	1	2	1	0.86*	Cs	55	133	$\frac{7}{2}$	2.57†
Na	11	23	$\frac{3}{2}$	2.22†	Hg	80	199	$\frac{1}{2}$	0.5
Al	23	27	$\frac{5}{2}$	3.6‡	Hg	80	201	$\frac{3}{2}$	-0.6
Rb	37	85	$\frac{5}{2}$	1.3	Tl	81	203, 205	$\frac{1}{2}$	1.6 ±
Rb	37	87	$\frac{3}{2}$	2.8	Bi	83	209	$\frac{9}{2}$	3.8 ±

* KELLOGG, RABI, and RAMSEY, *Phys. Rev.*, vol. 56, p. 728 (1929).

† KUSCH, MILLMAN, and RABI, *Phys. Rev.*, vol. 55, p. 1176 (1939).

‡ MILLMAN and KUSCH, *Phys. Rev.*, vol. 56, p. 303 (1939).

164. The Breadth of Spectral Lines.—No spectral line as observed is perfectly sharp, no matter how great the resolving power of the spectrometer. A line devoid of structure appears densest in the center and fades out symmetrically on the edges. It is of interest to inquire as to the causes of this distribution of the energy within a line.



FIG. 125.—Illustrating definition of Δ , the width of a line.

For convenience in comparing lines, the width Δ of a line is defined to be the distance between two points, one on each side of the center, at which the intensity is half as great as it is in the center of the line (Fig. 125). (This is sometimes called the half-intensity width.) The principal causes of line broadening are the following:

(a) *Doppler effect.*—The observed frequency of a spectral line may be slightly changed by the motion of the radiating atom in the line of sight, owing to Doppler's principle, the apparent frequency *increasing* if the motion is *toward* the observer and *decreasing* if the motion is *away* from the observer. Only for those atoms which have no component

¹ BACK and WULFF, *Zeits. f. Physik*, vol. 66, p. 31 (1930).

² CONDON and SHORTLEY, "Theory of Atomic Spectra," p. 424, 1935.

of velocity in the direction of the observer will the observed frequency of the emitted light be equal to the actual frequency. As is well known, the measurement of the radial velocity of stars is based on this principle.

In a luminous gas, such as the mercury vapor in the mercury arc lamp, the atoms are moving with a maxwellian distribution of velocity, the average velocity increasing with temperature. It is obvious, therefore, that a spectral line emitted by a gas must comprise, as observed, a *range* of frequencies *symmetrically* distributed about the frequency emitted by the atom when at rest and, further, that this range should increase with increasing temperature. The distribution of intensity throughout the line is determined by Maxwell's distribution of velocities. According to Rayleigh,¹ the brightness of the line at a distance $\pm \Delta \nu$ wave numbers from the center is proportional to

$$e^{-k\Delta\nu^2}$$

where k is a constant which depends on the temperature and on the mass of the atom, and e is the Napierian base of logarithms. From this formula it follows that the width Δ defined as above, if it is due entirely to the Doppler effect, should be

$$\Delta = 0.72 \times 10^{-6} \lambda \sqrt{\frac{T}{M}}$$

wave-length units, where T is the absolute temperature and M is the atomic or molecular weight of the radiating atom or molecule.²

Careful studies of the breadth of lines in the spectra of the rare gases were made by Fabry and Buisson,³ using an interferometer method. They confirmed the formula for Δ as written above. At liquid-air temperatures the breadth of the krypton line $\lambda = 5,570 \text{ \AA}$ was found to be only 0.006 \AA , practically all of which could be ascribed to the Doppler effect resulting from thermal agitation at that temperature.

(b) *Pressure*.—Using the light from a single unresolved spectral line, it is possible, under favorable conditions, to produce interference fringes when the difference in the path of the two beams is as much as several hundred thousand wave lengths. In terms of classical theory, this fact was interpreted to mean that the wave train sent out by any particular atom is continuous, *i.e.*, is without change of phase, for at least that number of vibrations. In order that the atom may emit wave trains of this length, it must be "free from interruptions" for a

¹ RAYLEIGH, *Phil. Mag.*, vol. 27, p. 298 (1889).

² Cf. WHITE, H. E., "Introduction to Atomic Spectra," p. 419, 1935.

³ FABRY and BUISSON, *J. de Physique*, vol. 2, p. 442 (1912).

corresponding period of time. In the terminology of the kinetic theory of gases, this means that the *mean free time* between collisions with other atoms must, on the average, exceed the time required to emit a complete wave train, since it may be assumed that a collision with another atom would cause either a change of phase or excessive damping or other disturbance. Analogous conclusions are deduced, although as a result of rather different reasoning, from wave mechanics.

Now, collisions between atoms become more frequent the higher the pressure of the gas for a given temperature. The higher the pressure, the shorter, therefore, should be the wave trains and the more frequent the abrupt changes of phase. An increase of temperature, at given density, also increases the collision rate and the pressure. Thus in general, at higher gas pressure, not only should there be a broadening of the line due to the Doppler effect but also an additional broadening due to the increasing frequency of phase changes resulting from collisions. Michelson confirmed this by showing, by measurements with the interferometer, that below a pressure of the order of a millimeter, the breadth of the hydrogen line $\lambda = 6563 \text{ \AA}$ is almost entirely due to the Doppler effect; but that at higher pressures the line becomes considerably broader.

The effect is enhanced by direct disturbances of the energy levels, or of the radiation process itself, when another molecule comes close to the radiating atom.

(c) *Natural Line Breadth*.—According to wave-mechanical theory, a line ought also to exhibit a small “natural” width even when emitted by an atom at rest. As an analogy, a classical oscillator, radiating energy, would decrease continually in amplitude; it would emit, therefore, a damped wave train of finite effective length. It can be shown that a damped train of sine waves is equivalent to the superposition of a large number of perfectly regular trains of great length, with slightly differing frequencies. Such radiation, observed as a spectral line, would, therefore, be broadened slightly. In the visible region, the natural line breadth is mostly much less than 0.001 \AA and so is not detectable. It is easily observable, on the other hand, in the case of X-rays.

(d) *Incipient Stark Effect*.—In the spectra emitted by discharge tubes, a common cause of broadening is the production of a small Stark effect by strong electric fields, which are not great enough to produce an observable splitting of the line but are sufficient to make it appear perceptibly broader than it otherwise would.

Thus there are many very different factors that may be responsible for the broadening of spectral lines.

165. Molecular Spectra.—In preceding sections of this chapter we have considered only spectra emitted by *atoms*. There are other spectra in vast variety which are believed on good evidence to be emitted by *molecules* containing 2 or more atoms.

Thus, in the visible part of the spectrum emitted by a discharge tube containing hydrogen, only three or four lines belonging to the Balmer series are emitted by free atoms, which have been produced by the dissociation of molecules. Many other lines are observed, however, mostly fainter, which are ascribed to emission by the undissociated molecules themselves. Again, if one looks at the spectrum of the carbon arc with a spectroscope of moderate resolving power, one will observe, at the extreme (violet) edge of the visible part of the spectrum, "bands," or very broad lines, which are sharply defined and brightest on the long-wave-length edge and which fade out gradually toward shorter wave lengths. With higher resolving power, these "bands" are seen to be composed of a large number of lines which are crowded together at the long-wave-length edge, called the "head" of the band, and are separated farther and farther toward the short-wave-length side, the lines, however, being so close together as to appear, under low resolving power, like a *continuous* spectrum. These bands are ascribed to molecules of cyanogen, CN. There are also many groups or bands of lines in the infrared which have been found to have a molecular origin.

In the space available here we can discuss only a few of the features of molecular spectra, referring the student to other books or to the literature for further information.¹

In general, the spectrum emitted by any given kind of molecule can be divided into *three spectral ranges* which correspond to different types of transition between molecular quantum states. The principal exception is that the spectrum of some molecules is confined to only one of these ranges. Simple reasoning in terms of familiar classical assumptions concerning molecular structure leads us to expect such a feature in molecular spectra, and the reasoning needs only to be translated into wave-mechanical terms in order to constitute a correct theoretical approach to the subject.

(a) *Rotation Spectra.*—Suppose a molecule were a rigid structure but contained electrical charges so disposed that the molecule possessed an electric moment. If such a molecule were to rotate, according to classical theory, it would emit radiation for essentially the same reason

¹ HERZBERG, G., "Molecular Spectra and Molecular Structure," 1939; RUARK and UREY, "Atoms, Molecules and Quanta," WEIZEL, W., "Bandenspektren," Akademische Verlagsgesellschaft m.b.H., Leipzig, 1931.

that an electron revolving in a circle would radiate. It is readily seen that the radiation would consist of sine waves having a single frequency, *viz.*, the frequency of rotation. Conversely, radiation falling upon such a molecule would tend to set it into rotation, energy of the radiation being at the same time absorbed.

A few spectral lines corresponding to this simple picture have been observed in the far infrared and are said to constitute "rotation spectra."

(b) *Vibration-rotation Spectra.*—If the molecule were not rigid but contained atoms capable of vibration under elastic forces about equilibrium positions, and if the chemical binding were of the ionic type (Sec. 134), so that some atoms contained an excess of positive charge and others an excess of negative, then according to classical theory radiation would be emitted by the vibrating atoms as they move back and forth. Unless the molecule were at the same time rotating, the frequency emitted would be that of the atomic vibration. If the molecule were rotating, however, the emitted line would be divided into two lines having frequencies respectively greater or less than the frequency of the atomic vibration, in essentially the same way as, in the classical theory of the Zeeman effect (Sec. 43), the precession caused by a magnetic field modifies the frequencies emitted by a vibrating electron.

Furthermore, it would be anticipated that, if the amplitude of vibration became large, the atomic vibrations, although still periodic, would no longer be simple harmonic. This is true even in the familiar example of the vibrations of a pendulum. The radiation emitted could then be resolved by Fourier analysis into wave trains with frequencies representing the fundamental and the harmonic overtones of the atomic vibrations. Each of these separate frequencies would then be split up further by rotation of the molecule.

Many spectra corresponding roughly to this classical picture are known in the infrared and are called "vibration-rotation spectra."

(c) *Electronic Spectra.*—Finally, according to classical ideas, an electron in the molecule might vibrate by itself and so radiate. The emitted radiation would be affected, however, both by the vibrations of the atoms in the molecule and by the rotation of the molecule as a whole. It would probably be one of the outer electrons that radiated in the optical region of the spectrum, and its frequency would be much affected by the instantaneous position and motion of the nuclei. The rotation of the molecule would then tend to split up the emitted lines as in the emission of the vibration-rotation spectrum. Molecular bands in the visible and ultraviolet, such as the cyanogen bands

described above, are believed to correspond roughly to this third classical picture.

The three types of molecular spectra thus characterized will be taken up individually for a brief discussion.

166. Rotation Spectra.—As a simple model to illustrate certain features of the behavior of actual molecules, we may imagine a molecule to consist of several mass points held rigidly at fixed distances from each other. The quantum states for such a molecule, according to wave mechanics, would be characterized by fixed values for the angular momentum, in the same way as are the states of an atom (Sec. 140). The corresponding quantum number J , however, is confined here to integral values (zero included). The discussion will be restricted hereafter almost entirely to *diatomic molecules*. If there are only two mass points in the molecule, the line joining them is an axis of symmetry, and only rotation about an axis perpendicular to this line has meaning; furthermore, the moment of inertia about all such perpendicular axes will have the same value.

The relation between angular momentum and energy is found to be the same according to wave mechanics as in classical theory. For the angular momentum G and energy W , we shall have, therefore, in terms of the angular velocity ω and moment of inertia I ,

$$G = I\omega, \quad W = \frac{1}{2}I\omega^2, \quad \therefore W = \frac{G^2}{2I}.$$

Inserting here the wave-mechanical value of G^2 (Sec. 140), we have

$$G^2 = \frac{J(J+1)h^2}{4\pi^2},$$

$$W = J(J+1) \frac{h^2}{8\pi^2 I}. \quad (229)$$

Such a molecule can radiate only if it possesses an electrical moment. Such a moment will exist, for example, if one mass point has associated with it a positive charge and the other an equal negative charge. The selection rule for J is then found to be the same as that for the quantum number l of a single electron in a central field:

$$\Delta J = \pm 1. \quad (230)$$

Since in the present case W and J increase or decrease together, $\Delta J = -1$ will correspond to emission of energy and $\Delta J = +1$ to absorption. Thus, in a transition from state J to state $J-1$, the emitted frequency will be, since

$$J(J+1) - (J-1)(J-1+1) = 2J,$$

$$\nu = \frac{\Delta W}{h} = 2BJ, \quad B = \frac{h}{8\pi^2 I}. \quad (231a,b)$$

Hence a molecule of the type under consideration will emit a spectrum consisting of *equally spaced lines, with frequencies equal to a multiple of a fixed number, B*. For an emission line, J refers to the initial state for the molecular transition; for an absorption line, to the final state. The corresponding type of energy-level diagram is illustrated in

Fig. 126. The arrows pointing downward refer to transitions for emission, those pointing upward to transitions for absorption.

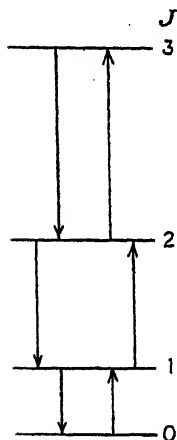


FIG. 126.—
Energy-level diagram for a rotation spectrum.

We may form some idea of the spectral region in which such lines may be expected to occur by estimating the probable magnitude of I from other considerations. The molecule HCl , for example, might be expected to have an electric moment, since it is strongly ionized in solution. In the formula $I = \Sigma mr^2$, in which r denotes distance from the center of mass of the molecule, the Cl atom will contribute little because it is so close to the center of mass. The mass of a hydrogen atom will be approximately the reciprocal of the Avogadro number (Sec. 85) or the number of atoms in a gram-atom. Hence, if we assume $r = 10^{-8}$ cm. for the H atom, we have

$$I = 10^{-16} / (6 \times 10^{23}) = 1.7 \times 10^{-40} \text{ c.g.s. unit.}$$

Thus, for the constant B in Eq. (231a), if we also divide by c to obtain it in wave-number units, we find

$$B = \frac{6.6 \times 10^{-27}}{8\pi^2 \times 1.7 \times 10^{-40} \times 3 \times 10^{10}} = 17 \text{ cm.}^{-1},$$

roughly. Thus, for $J = 10$, we find $\bar{\nu} = 2BJ = 340 \text{ cm.}^{-1}$, or $\lambda = 0.029 \text{ mm.} = 29\mu$.

Now all of the hydrogen halides in the gaseous state show broad absorption lines in the far infrared which are very nearly equally spaced and whose wave numbers are very nearly multiples of a constant quantity. These lines are believed to result from transitions in which the rotational state of the molecule alone changes, very nearly according to the simple theory just described. Thus for HCl , Czerny¹ found

¹ CZERNY, *Zeits. f. Physik*, vol. 34, p. 227 (1925).

a series of absorption maxima at wave lengths ranging from 120μ to 44μ . The corresponding wave numbers are listed as $\tilde{\nu}_{\text{obs}}$ in Table XI.

TABLE XI.—ABSORPTION SPECTRUM OF HCl IN THE FAR INFRARED

J	$\tilde{\nu}_{\text{obs}}$	$\Delta\tilde{\nu}$	$\tilde{\nu}_{\text{calc}} = fJ - gJ^2$
4	83.03		83.06
5	(41.27)	103.75
6	124.30		124.39
7	145.03	20.73	144.98
8	165.51	20.48	165.50
9	185.86	20.35	185.94
10	206.38	20.52	206.30
11	226.50	20.12	226.55

Under the heading J is given the larger of the two values of J assigned to each transition. We note that the spacing of the maxima is almost uniform but shows a slight trend. In the last column of the table are shown values of $\tilde{\nu}$ calculated from the empirical formula, $\tilde{\nu} = fJ - gJ^2$, with $f = 20.79$, $g = 0.0016$; according to this formula

$$\Delta\tilde{\nu} = f - g[J^2 - (J - 1)^2] = f - (3J^2 - 3J + 1)g \text{ cm.}^{-1}.$$

If we have sufficient confidence in the theoretical interpretation, such data can be used as a basis for the calculation of molecular moments of inertia. Substituting $B = cf/2 = 20.79 c/2$ in Eq. (231b), we find for the moment of inertia of the HCl molecule, in a state of minimum rotation, $I = 2.45 \times 10^{-40}$ gram cm.². This value agrees satisfactorily with our rough estimate, 1.7×10^{-40} . The slight decrease in $\Delta\tilde{\nu}$ as J increases, and hence also in the apparent value of B , implying an increase in I , is what would be expected if the atoms were not tightly bound together but became slightly pulled apart by centrifugal action as the speed of rotation increases.

Thus, wave mechanics accounts satisfactorily for the rotation type of molecular spectrum.

167. Vibration-rotation Spectra. (a) *Approximate Theory of a Vibrating Diatomic Molecule.*—To obtain a simple model of a diatomic molecule in which atomic vibrations can occur, suppose the atoms are themselves point masses but that, instead of being rigidly bound together, they are held by a force that varies with the distance between them. Let this force correspond to a potential energy V , which, plotted as a function of the distance r between atomic centers, is represented by a curve of the type of that marked V in Fig. 127. The potential has been arbitrarily taken to be zero at $r = \infty$. The force

exerted by either atom on the other will be proportional to the slope of this curve; it is represented by the curve marked F in Fig. 127. From the point r_0 at which V has its minimum value outward, the force is attractive; for $r < r_0$, it is repulsive, rising rapidly to high values. Under the influence of such a force, according to classical theory, the atoms could be at rest and in equilibrium at the distance r_0 . If disturbed moderately, they would vibrate about this point; if given kinetic energy exceeding $-V_0$, however, where V_0 is the value of V at $r = r_0$, they would fly apart entirely, *i.e.*, dissociation of the molecule would ensue. Such a picture corresponds to the observed properties of molecules, and it is also suggested by wave-mechanical theory.

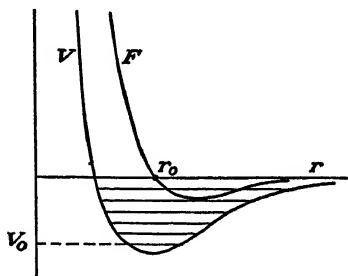


FIG. 127.—Illustration of the mutual potential energy V and force F , measured positively as a repulsion, between 2 atoms in a diatomic molecule.

$h\nu/2$ of a harmonic oscillator. Thus the interatomic distance is not quite fixed even when the molecule is in this state, although the most probable value of the interatomic distance will be close to r_0 . Let us number the quantum states in the order of increasing energy, denoting the number of a state by $v \geq 0$. Let the corresponding energy be denoted by W_v . The values of W_v are all negative, zero energy belonging to a state in which the atoms are at rest at infinity. Thus, $-W_v$ represents the energy of dissociation of the molecule when it is in state number v and not rotating. The total number of the discrete states may be finite or infinite, depending upon the form of the potential curve.

The wave functions and energies for the first few states should resemble those for a harmonic oscillator. For, if we expand V in a Taylor series about $r = r_0$ and note that $dV/dr = 0$ at $r = r_0$, we obtain

$$V = V_0 + \frac{1}{2} \left(\frac{d^2V}{dr^2} \right)_{r=r_0} (r - r_0)^2 + \frac{1}{6} \left(\frac{d^3V}{dr^3} \right)_{r=r_0} (r - r_0)^3 - \dots,$$

and the first two terms of this series represent a potential function

of the same type as that for a harmonic oscillator (Sec. 123).¹ The wave functions and energies for the higher states, however, must depart considerably from those for an oscillator, owing to the influence of the remaining terms of the series. Thus the selection rule for the harmonic oscillator, $\Delta v = \pm 1$, cannot be expected to hold, in general, for a diatomic molecule, even $\Delta v = 0$ being possible. The probability of a jump may be expected, however, to fall off rather rapidly as Δv increases.

To obtain quantum states for the whole molecule, allowance must then be made for rotation of the line joining the atoms. If we suppose that, to a sufficient approximation, the energies of rotation and of vibration are additive, we may write for the total energy, by Eq. (229) above,

$$W = W_v + J(J + 1)Bh, \quad B = \frac{h}{8\pi^2 I} \quad (232a, b)$$

The various values of W_v are called *vibrational* levels of the molecule. The energies represented by both terms on the right in Eq. (232a) are called *vibration-rotation* levels, or simply *rotational* levels. In actual cases Bh is usually very small relative to the difference between successive values of W_v ; hence, the rotational levels belonging to each vibrational level form a closely spaced group. The general arrangement of the rotation-vibration levels is illustrated in Fig. 128, in which, however, the relative spacing of the rotational levels is enormously exaggerated.

Because of the relative smallness of B , all lines arising from transitions between two given vibrational levels will lie close together; they are said to constitute a *band*, because with low resolving power they appear as a continuous streak in the spectrum. Bands arising from transitions in which only the vibrational and perhaps rotational energies of the molecule change are called *vibration-rotation* bands.

¹ The wave functions for the first few states should be practically zero for values of $(r - r_0)$ that are large enough to give appreciable magnitude to terms of the series beyond the first two.

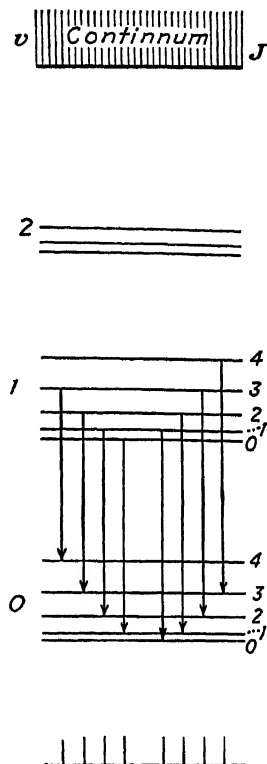


FIG. 128.—Diagram illustrating the energy levels for a rigid diatomic molecule (the relative magnitude of the rotational separations being enormously exaggerated), and a few of the transitions for the first band.

For a molecule composed of two point masses, the selection rule for J is found to be, as for the pure rotation lines,

$$\Delta J = \pm 1.$$

For many actual molecules, however, $\Delta J = 0$ is also allowed. A few transitions allowed by the rule $\Delta J = \pm 1$ are indicated in Fig. 128.

From the approximate expression for the energy in Eq. (232a), we find, since

$$J(J+1) - (J-1)(J-1+1) = 2J,$$

for the frequencies emitted in a transition between levels $v', J-1$ and $v''J$, or between $v'J$ and $v'', J-1$,

$$v', J-1; v''J: \quad \nu = \nu_{v''v'} - 2BJ, \quad J = 1, 2, 3, \dots; \quad (233a)$$

$$v'J; v'', J-1: \quad \nu = \nu_{v''v'} + 2BJ, \quad J = 1, 2, 3, \dots; \quad (233b)$$

$$\nu_{v''v'} = \frac{W_{v'} - W_{v''}}{h}. \quad (233c)$$

It is assumed here that $W_{v'} > W_{v''}$.*

According to Eqs. (233a,b), the lines in a given vibration-rotation band should be equally spaced in frequency, with a separation $2B$. It is clear, however, that the value $\nu = \nu_{v''v'}$ cannot occur, so that there is no line corresponding to pure vibration. The nearest frequencies to this are $\nu = \nu_{v''v'} + 2B$ and $\nu_{v''v'} - 2B$. The central line of the band should thus appear to be missing.

(b) *Observed Absorption Bands of Vibration-rotational Type.*—Absorption bands answering the foregoing description were observed long ago in the near infrared regions of the spectra of the hydrogen halides. The band of HCl at 3.5μ is shown, as observed by Imes,¹ in Fig. 129, percent of absorption being plotted against wave length. In order to obtain this curve, Imes passed the radiation from a Nernst glower through 15 cm. of HCl gas at atmospheric pressure and then through a spectrometer containing a prism of rock salt, and measured the transmitted radiation with a thermopile and a galvanometer. The band in question had been observed by Burmeister in 1913 as a continuous band with a double top and was thought to be capable of explanation in terms of classical theory. The molecules were assumed to possess a maxwellian distribution of rotational velocities, so that, according to classical theory, the splitting described in Sec. 164(a)

* In absorption, therefore, v' refers to the final state into which the molecule is raised by the absorption of a quantum. In considering the spectrum, however, it is simplest to think only of numerical energy differences, without bothering to remember which state is the initial or final one in the radiative process.

¹ IMES, *Astrophys. J.*, vol. 50, p. 251 (1919).

would be expected to spread the vibrational line out into a broad, continuous band. When later workers observed a succession of maxima in the band, it was realized that an explanation could only be found in terms of a discrete set of rotational quantum states.

The anticipated absence of the central line in Fig. 129 is very evident. The lines show some approach to equal spacing, but considerable variation occurs; this is well brought out by the following table of frequency differences, in which those values of J are primed that refer to transitions in which $\Delta v = v'' - v'$ and ΔJ have opposite

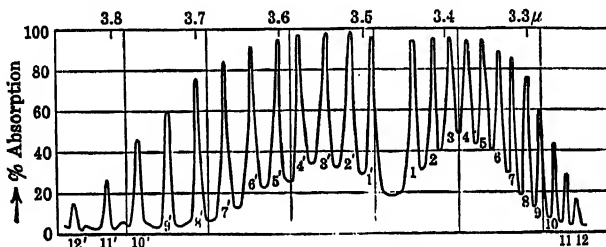


FIG. 129.—The principal absorption band of HCl in the near infrared (after E. S. Imcs¹). The numbers on the lines give the larger J for the transition, those for which ΔJ is opposite in sign to Δv being primed. (Reprinted by courtesy of Prentice-Hall Inc., New York; from G. Herzberg, "Molecular Spectra and Molecular Structure," 1939.)

signs, so that the change in rotational energy is opposite in sign to the change in vibrational energy and the frequency is less than $\nu_{v''v'}$:

$$v', J - 1; v''J$$

J	12'	11'	10'	9'	8'	7'	6'	5'	4'	3'	2'	1'
$\Delta\bar{\nu}$	26.62	26.15	25.91	25.03	25.30	23.07	23.85	23.04	22.40	22.48	21.75	

$$v''J; v', J - 1$$

J	(1')	1	2	3	4	5	6	7	8	9	10	11	12
$\Delta\bar{\nu}$	41.60	20.24	18.35	18.02	17.38	17.62	16.30	14.92	15.89	14.47	13.27	12.85	

(c) *The More Accurate Empirical Theory of Vibration-rotation Bands.*—The variation in $\Delta\bar{\nu}$ as observed indicates that the theory developed above, based upon a molecule composed of two point-mass atoms, is too simple and requires modification. Even on classical assumptions, complications are to be expected in an actual atom. When the molecule is in rotation, the position of relative equilibrium for the 2 atoms should shift outward from r_0 in Fig. 127, because of centrifugal action; the result would be an increase in the moment of inertia I . An effect of the spreading upon the vibrational motion of the molecules is also to be anticipated.

The true formula for the energy levels must, therefore, be more complicated than that expressed in Eq. (232a), although the quantum

¹ Loc. cit.

numbers v and J can undoubtedly be retained as labels. Various empirical equations have been proposed from time to time, alongside efforts to improve the treatment based on quantum theory. A form of equation equivalent to the following has become accepted as standard, all terms being expressed in cm.^{-1} :

$$\bar{W} = \bar{W}_v + B_v J(J+1) - D_v J^2(J+1)^2 \dots \quad (234)$$

The coefficients B_v and D_v vary somewhat with v and are often written as power series in this variable:

$$B_v = B_0 - \alpha v + \gamma v^2, \quad D_v = D_0 + \beta v^2 \dots$$

In these series, wave-mechanical considerations indicate that B_0 and D_0 should be positive, and that D_v as a whole should be comparatively small. If we omit the D_v term entirely, we obtain, in place of (233a) and (233b), changing to wave numbers instead of frequencies:

$$v', J-1; v'', J: \quad \bar{\nu} = \bar{\nu}_{v''v'} - (B_{v'} + B_{v''})J + (B_{v'} - B_{v''})J^2 \dots \quad (235a)$$

$$v', J; v'', J-1: \quad \bar{\nu} = \bar{\nu}_{v''v'} + (B_{v'} + B_{v''})J + (B_{v'} - B_{v''})J^2 \dots \quad (235b)$$

These equations yield the lines in order of increasing frequency if we take decreasing values of J in the first equation and increasing values in the second. For $\Delta\bar{\nu}$ between successive lines, we find, denoting by J_m the larger of the two values of J for each transition:

$$\begin{aligned} v', J-1; v'', J: \quad \Delta\bar{\nu} &= B_{v'} + B_{v''} - (B_{v'} - B_{v''})(2J_m - 1) \dots, \\ v', J; v'', J-1: \quad \Delta\bar{\nu} &= B_{v'} + B_{v''} + (B_{v'} - B_{v''})(2J_m - 1) \dots. \end{aligned}$$

In view of the direction in which J changes throughout the band, these equations imply that, if $B_{v'} < B_{v''}$, $\Delta\bar{\nu}$ will decrease from one line to the next, as it actually does. The observed rate of decrease, however, is greater at higher than at lower frequencies. This is easily seen to be accounted for by the term in D_v in Eq. (234), provided $D_{v'}$ and $D_{v''}$ are both positive. The term in D_v represents essentially the stretching effect due to centrifugal action, which was referred to above; this stretching, by increasing the moment of inertia I , should decrease the energy, as is evident from the approximate equations (232a,b); hence, in Eq. (234), D_v should be positive.

(d) *Relation between Rotation and Vibration-rotation Bands.*—It is obvious that there must exist a close relation between the vibration-rotation bands and the pure rotation spectrum of a given substance; indeed, the rotation spectrum is simply a band arising from transitions in which $\Delta v = 0$. If we put $\nu_{v''v'} = 0$ in Eq. (233b), we obtain

$\nu = 2BJ$, in agreement with (231a) in the last section. Thus, according to the simplest type of theory, the *separations of the lines should be the same in the rotation and in the vibration-rotation parts* of the spectrum, being equal to $2B$ in both cases. This conclusion is easily tested. According to (233a) and (233b), the separation of the innermost two lines ($J = 1$) in a vibration-rotation band is $4B$. In the HCl band just described, this separation is 41.60 cm.^{-1} . Half of this, or 20.8, agrees exactly with the value obtained for the spacing of the lines at $J = 0$ from the empirical formula cited in the last section for the rotation spectrum of HCl, *viz.*, $\Delta\tilde{\nu} = f = 20.79$. The comparison will not be pressed more closely here, because of insufficient accuracy in the data for the vibration-rotation band.

(e) *Bands of Higher Order, and the Heat of Dissociation.*—Besides the relation between the two kinds of bands, another theoretical point that can be tested is the prediction that, although transitions for $\Delta\nu > 1$ are to be expected, the resulting bands should be relatively weak. If the vibrational levels were equally spaced, as for the harmonic oscillator [Eq. (184a,b) in Sec. 123], the vibrational frequencies of the bands for $\Delta\nu \geq 1$ would be proportional to 1, 2, 3, . . . , just as they are for the harmonic overtones of a classical vibrating system. Thus, if we assume that the HCl band at 3.46μ arises from a transition between the lowest two vibrational levels ($v'' = 0, v' = 1$), we should expect to find bands diminishing progressively in intensity at about $1.73\mu, 1.15, 0.865$, etc. Actually, by observing the absorption through very thick layers of HCl gas, bands have been observed at $1.76, 1.20$, and 0.916μ , the last mentioned being 10,000 times weaker than the 3.46 band.¹

The occurrence of these "harmonic" bands with the expected variation in strength, and the agreement of the spacing in all bands, including the rotation spectrum, with theoretical expectations, constitute strong evidence that our interpretation of these infrared absorption spectra is essentially correct. A further check is furnished by a connection with the *heat of dissociation* of the molecule. The heat of dissociation represents the energy that must be added to a molecule in its lowest quantum state in order to separate the atoms and leave them at rest an infinite distance apart. Obviously, the energy of the molecule in any quantum state must be less than the energy of the separated atoms, else the molecule would dissociate spontaneously. Hence, the quantity $h\nu$ for any molecular line, representing the difference in energy between two molecular states, must be less than the

¹ HERZBERG and SPINKS, *Zeits. f. Physik*, vol. 89, p. 474 (1934); CLEAVES and EDWARDS, *Phys. Rev.*, vol. 48, p. 850 (1935).

heat of dissociation. For the HCl band at $\lambda = 0.916\mu$,

$$h\nu = 6.6 \times 10^{-27} \times \frac{3 \times 10^{10}}{0.916 \times 10^{-4}} = 2.2 \times 10^{-12} \text{ ergs.}$$

In the "International Critical Tables,"¹ the heat of dissociation of a gram-molecule of HCl, for dissociation into H_2 and Cl_2 , is given as 92 kilojoules, whereas the heats of combination of a gram-atom of H or Cl into H_2 or Cl_2 are, respectively, 211 and 120.3 kilojoules. Hence to dissociate 1 molecule of HCl into H and Cl requires

$$(92 + 211 + 120.3) \times \frac{10^{10}}{6.02 \times 10^{23}} = 7.0 \times 10^{-12} \text{ erg.}$$

This is more than three times the value of $h\nu$ as just calculated for the highest-frequency vibration band of HCl that has been observed, so that the theoretical expectation is confirmed.

(f) *Effect of Thermal Agitation upon Infrared Bands.*—No absorption bands are observed for which $\nu'' > 0$ in the spectrum of HCl. This is explained as a consequence of the wide spacing of the vibrational levels relative to the quantity kT . Molecules are seldom thrown by thermal agitation into a vibrational level which can serve as the initial level for an absorption band with $\nu'' > 0$. Even for $\lambda = 3.5\mu$, $h\nu = 5.7 \times 10^{-13}$ erg; whereas at $T = 288^\circ\text{K.}$ (15°C.),

$$kT = 288 \times 1.38 \times 10^{-16} = 4 \times 10^{-14} \text{ erg.}$$

If W_0, W_1 are the energies of the two states involved in the production of $\lambda = 3.5\mu$, the ratio of the number of atoms in the upper of these two states to that in the lower will be the ratio of their Boltzmann factors or, according to Eq. (148a) in Sec. 110,

$$\frac{n_1}{n_0} = e^{-(W_1 - W_0)/kT} = e^{-h\nu/kT} = e^{-6.74} = 10^{-6.2}.$$

This is so small that practically all molecules of HCl will be in their lowest vibrational states ($\nu = 0$) at ordinary temperatures. Hence only the absorption band for which $\nu'' = 0$ can be observed; and the observed rotation spectrum is the $0-0$ vibration-rotation band ($\nu = 0$ to $\nu = 0$).

Among the various *rotational* levels, on the other hand, a wide distribution of the molecules occurs even at room temperature. For the experimental value of B , $20.6/2 \text{ cm.}^{-1}$, makes Bh equal to

$$6.6 \times 10^{-27} \times 3 \times 10^{10} \times \frac{20.6}{2} = 2.0 \times 10^{-16} \text{ erg.}$$

¹ Vol. V, p. 176.

Thus, for these states, if the energy of the lowest is taken as zero, $W = J(J + 1) \times 2 \times 10^{-15}$ erg, which is much smaller than kT . For $J = 10$, we find $e^{-W/kT} = e^{-5.5} = 0.0041$. Furthermore, there are $2J + 1$ fundamental states for each J (with $M = -J, \dots, J$), so that even states with J as large as 10 will occur

$$(2 \times 10 + 1) \times 0.0041 = 0.08$$

times as often as states with $J = 0$. This result agrees nicely with the number of lines that are actually observed in the HCl band at 3.5μ .

In some other substances, such as I_2 , an appreciable fraction of the molecules are normally in higher vibrational states, and other bands can occur in absorption or in the rotation spectrum.

(g) *Infrared Bands of Other Types of Molecules.*—Vibration-rotation spectra are known for many other molecules than those which have been mentioned. Their structure is often more complex than that described above; in many cases transitions for $\Delta J = 0$ are permitted, so that an additional sequence of lines occurs and there is no gap in the center of the band. Thus the absorption spectrum of CO_2 shows many vibration-rotation bands from 1.46μ to 15.05μ , and that of water vapor shows many from 0.69μ to $6.26\mu^*$; these bands, together with the rotation lines of water vapor, are responsible for the marked absorption of the earth's atmosphere in the infrared.

In order to exhibit vibration-rotation and rotation spectra of appreciable intensity, a molecule must possess an electrical moment. *Homonuclear* diatomic molecules, such as O_2 , N_2 , H_2 , Cl_2 , possess no moments and, hence, have no spectra of these two types. Gases composed of such molecules are entirely transparent in the infrared region.

168. General Theory of Molecular Quantum States.—Up to this point, the *electrons* in the molecule have been ignored. Actually, the wave function for a molecule must contain the coordinates and spins of all of the electrons as well as the coordinates of the nuclei. When atoms unite into a molecule, the inner electrons in each atom can be regarded as remaining associated, more or less, with the nucleus of that atom; but the outer electrons come to belong to the molecule as a whole rather than to any individual nucleus.

(a) *Approximate Separation of Electronic and Nuclear Motions.*—In many cases, especially for the diatomic molecules, the wave function can be separated approximately into two factors, of which one has reference to the electrons, whereas the second factor represents

* WEIZEL, *loc. cit.*

vibrations of the nuclei and rotation of the molecule as a whole. The first factor then represents the electrons as being in a certain *electronic quantum state*. This electronic factor is analogous to the wave function for the electrons in an atom. Both the energy and the wave function of the electronic state, however, will depend upon the relative positions of the nuclei. In this respect the situation in a molecule is quite different from that in the atom. The mutual electrostatic energy of the nuclei is commonly included in the energy of the electronic state. The electronic energy as so defined possesses a minimum value for certain relative positions of the nuclei; if they move closer together, or move farther apart, the electronic energy rises and, hence, so does the energy of the molecule. Thus the energy of the electronic state is equivalent to a potential energy tending to hold the atoms together in definite relative positions of equilibrium. It is in this way that an interatomic potential energy arises in a diatomic molecule such as was sketched in Fig. 127.

In this same approximation, the energy W of the molecule can be written as the sum of two parts, a negative part W_e , representing roughly the average of the electronic energy, including also the electrostatic energy of the nuclei, and a much smaller positive part W_{rv} , which is associated with vibration of the atoms relative to each other and with rotation of the molecule as a whole. Thus

$$W = W_e + W_{rv}. \quad (236)$$

For a diatomic molecule, the energy W_{rv} is what was denoted by W in Sec. 167, as in Eq. (232a) or (234).

For a detailed discussion of electronic states we must refer the student to other books,¹ but something may be said here by way of explanation of the notation that the student will meet. Only *diatomic* molecules will be discussed.

(b) *ΔS Coupling*.—As in the theory of atomic states, electronic spin may be treated in different ways according to circumstances. If the *electronic spin-orbit effects are small*, it is possible, as a first approximation, to assign a fixed value to the component of the orbital angular momentum of the electrons about the nuclear line, or line joining the two nuclei, which is an axis of symmetry. The magnitude of this momentum is denoted by $\Lambda h/2\pi$, where Λ is some positive integer or zero. The quantum number Λ corresponds to A or M_L in atomic theory, which refers to the component of the angular momentum about the axis of coordinates. In a diatomic molecule a unique choice of axis is supplied by the nuclear line. The *total* orbital momen-

¹ Cf. HERZBERG, "Molecular Spectra and Molecular Structure," Chaps. IV, V.

tum of the electrons cannot have a fixed value in a molecule; hence there can be no quantum number L . Spectroscopically, however, Λ plays a role somewhat analogous to L , since states corresponding to different values of Λ have different energies. Hence, in imitation of the atomic notation for LS coupling, states with $\Lambda = 0, 1, 2, 3, \dots$ are indicated by the letters $\Sigma, \Pi, \Delta, \Phi, \dots$, respectively.

The electronic spins are then combined, with the introduction of a quantum number S . The value of $2S + 1$ is often written as a superscript. Thus, we obtain a set of ΛS electronic states represented by such symbols as

$$^1\Sigma, ^3\Sigma, ^1\Pi, ^3\Pi, \dots$$

As with atoms, an important significance of the quantum numbers lies in the associated selection rules. There is a strong tendency for transitions to be limited to those for which

$$\Delta\Lambda = 0 \text{ or } \pm 1, \quad \Delta S = 0.$$

Instead of introducing next the electronic spin-orbit effects, however, as is done in atomic theory, we now turn to the consideration of the *nuclear motions*, the effect of which upon the energy we are assuming to be much larger than the electronic spin-orbit effects. Each electronic state of the molecule may be combined with any one of the vibrational states for the nuclei, which may be numbered off as before with a quantum number $v = 0, 1, 2, \dots$. The orbital angular momentum of the electrons is then added vectorially to the angular momentum of rotation of the nuclei about axes perpendicular to the nuclear line. This results in the introduction of a rotational quantum number K , which may have any integral value such that

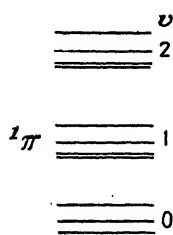
$$K \geq \Lambda.$$

The quantum number K has the significance that the square of the resultant orbital angular momentum of electrons and nuclei together has the value $K(K + 1)\hbar^2/4\pi^2$. The general selection rule for K is that

$$\Delta K = 0 \text{ or } \pm 1.$$

We obtain in this way a set of *electronic-vibration-rotation states* for the molecule numbered with quantum numbers $\Lambda S v K$. But the *electronic spin-orbit effect* has not yet been considered. In preparation for its introduction, we must now add the total orbital angular momentum of electrons and nuclei, represented by K , to the total angular momentum due to electronic spin, represented by S , thus obtaining a grand resultant angular momentum, for which we will

introduce the usual quantum numbers J and M . The possible values of J are integrally spaced from $K + S$ down to $|K - S|$. The effect of the electronic spin-orbit interaction is then, as in atomic LS coupling, to separate states having different values of J . Thus a *fine structure*, analogous to the LS multiplet structure, is introduced into the electronic-vibration-rotation levels. The superscript in such a symbol as ${}^3\Pi$ refers to the (normal) number of J levels in this fine structure.



In singlet levels, with $S = 0$, there is no fine structure, just as in atoms. For such levels, $J = K$, and the quantum number K need not be used. The HCl molecule, for example, is normally in a ${}^1\Sigma$ electronic state; hence, the discussion of the vibrational and rotational states of this molecule as given above was adequate.

(c) Ω Coupling.—When the electronic spin-orbit effects are not smaller than the effects of molecular rotation, other modes of approach by means of perturbation theory become appropriate. We shall mention only one of them, briefly. If the *spin-orbit* effect is actually *large*, we have the analog of jj coupling in atoms. In this case, the electronic orbital and spin momenta about the nuclear line are first added together, their magnitude being represented by a quantum number Ω . The values of Ω are integral or half-integral according as the number of electrons in the molecule is even or odd. Each electronic state, characterized by a value of Ω , is then combined with a vibrational state represented by quantum numbers v and J . Thus in this form of coupling, as when $S = 0$, the rotational levels correspond to the value of J .

FIG. 130.—Diagram illustrating general arrangement of molecular levels. To make the scale correct, the rotational levels (numbered by $K = J$) should be drawn much closer together, and the ${}^1\Pi$ and ${}^1\Sigma$ groups should lie much farther apart.

The *energy of a molecular state* thus depends, in the case of ΛS coupling upon five quantum numbers, $\Lambda S v K J$, or in the case of Ω coupling upon three, $\Omega v J$. Its variation with the various numbers presents, however, several different orders of magnitude. The variation with K or J is comparatively small. Spectral lines that differ only in the K values for the initial and final states lie close together and form a band, similar to the vibration-rotation bands that have been described. In such cases, if $J \neq K$, the energy varies still less with J , the effect of this variation being only to introduce a fine structure into the lines of the band. In the case of Ω coupling, on the other hand, the lines of a band arise from differences in

the values of J . The energy varies much more rapidly with v than it does with K or J . Each pair of values of v , one for the initial and one for the final state, gives rise to a possible band. The electronic states, finally, characterized by various values of Λ and S , or of Ω , are separated in energy by differences of the order of those between atomic LS terms. These general features of the array of molecular levels are illustrated for a simple case in Fig. 130.

169. Electronic Bands.—The most general type of transition between molecular states is one in which changes occur in the electronic state of the molecule as well as in its nuclear vibration-rotational state. The spectra hitherto discussed, of rotational or vibrational type, represent special cases in which the electronic state does not change. Such spectra constitute, however, only a small fraction of all known band spectra. When the electronic state does change in a transition, the resulting change in energy is usually so large that the band lies in the visible or ultraviolet region of the spectrum. Such bands may be called *electronic* bands.

In transitions characterized by a given pair of electronic states and by given values v' and v'' of v , representing a fixed pair of vibrational states, various changes of J (and of K , if $K \neq J$) may occur. The resulting lines form a single *band*. All of the bands due to transitions between a given pair of electronic states, for all possible values of v' and v'' , are said to form a *band system*. Because various electronic jumps are possible, the band spectrum of any molecule consists of many band systems.

The lines in a given electronic band are limited by the selection rules

$$\Delta K = 0 \text{ or } \pm 1, \quad \Delta J = 0 \text{ or } \pm 1.$$

Aside from the fine structure that exists when the rotational levels are numbered by the quantum number K and $K \neq J$, electronic bands are commonly more complicated than the simple type of vibration-rotation band described above because transitions for $\Delta K = 0$ (or $\Delta J = 0$) are allowed. Lines for which ΔK (or ΔJ if there is no fine structure) has the opposite sign to Δv are said to constitute the *P branch* of the band; those for $\Delta K = 0$ (or $\Delta J = 0$) constitute the *Q branch*; those for which ΔK (or ΔJ) is in the same direction as Δv form the *R branch*. The *Q branch* is frequently missing, however, for special reasons, *e.g.*, in all bands arising from a $\Sigma \rightarrow \Sigma$ electronic transition. Sometimes when $K \neq J$, what is regarded as a single band contains more than one branch of each type.

A good approximate expression for the energies of the molecular levels is often obtained if we employ an expression like that in Eq.

(234) for the vibration-rotation part. Omitting the small D_e term, we have then for the molecular energy, in cm^{-1} ,

$$\bar{W} = \bar{W}_e' + \bar{W}_e'' + B_e K(K+1) \cdots \quad (237)$$

Thus, for the levels belonging to two different electronic states, we may write

$$\begin{aligned}\bar{W}' &= \bar{W}_e' + \bar{W}_v' + B_v' K(K+1), \\ \bar{W}'' &= \bar{W}_e'' + \bar{W}_v'' + B_v'' K(K+1);\end{aligned}$$

and we then find for the three branches of the band, as a generalization of (235a) and (235b):

$$P: v', K-1; v'', K: \bar{\nu} = \bar{\nu}_e + \bar{\nu}_{v''} - (B_v' + B_v'')K + (B_v' - B_v'')K^2 \quad (238a)$$

$$Q: v'K, v''K: \bar{\nu} = \bar{\nu}_e + \bar{\nu}_{v''} + (B_v' - B_v'')K(K+1) \quad (238b)$$

$$R: v'K; v'', K-1: \bar{\nu} = \bar{\nu}_e + \bar{\nu}_{v''} + (B_v' + B_v'')K + (B_v' - B_v'')K^2 \quad (238c)$$

Here K refers in each case to the larger of the two values of K concerned in the transition. If the quantum number K does not exist for the levels in question, K is to be replaced by J in all equations from (237) to (238c). The symbol $\bar{\nu}_e = (\bar{W}_e' - \bar{W}_e'')/h$ and represents the frequency that would arise from the electronic transition alone; whereas $\bar{\nu}_{v''} = (\bar{W}_v' - \bar{W}_v'')/h$.

Since the quantities B_v' and B_v'' in Eqs. (238a) to (238c) refer to different electronic states, they may be expected to differ considerably. We may suppose the forces between the atoms to be quite different in the two cases; hence, their positions of equilibrium and the values of the moment of inertia of the molecule are also different. This is in contrast to the case of the vibration-rotation bands, where B_v varies only a little from one level to another. The quadratic terms in (238a,b,c), therefore, will soon make themselves felt. As K increases, the trend of $\bar{\nu}$ in one branch will soon be reversed, in the P branch if $B_v' > B_v''$, in the R branch if $B_v' < B_v''$. It is thus a general characteristic of electronic bands that one branch is folded back on itself and on top of the others. At the point in the spectrum where the branch turns back, the lines are crowded together, forming a *band head*. The band appears to shade away from the head; some bands are shaded in this way toward the red, some toward the violet. When a Q branch is present, it may form a second head, although such behavior is not obviously predicted by Eq. (238b).

A band of this type is often represented graphically by means of a Fortrat diagram, on which each line is represented by a point or circle, the ordinate representing K (or J) and the abscissa, $\bar{\nu}$.

Such a diagram is shown in Fig. 131 for the CuH band $\lambda 4,280$ A, which has no Q branch; a spectrogram of this band due to Professor R. Mecke is shown in Fig. 132. The band is shaded toward the red from a head which is plainly shown in the spectrogram. The electronic transition is $^1\Sigma \rightarrow ^1\Sigma$.

The number of different bands in a band system, all arising from the same electronic transition but with various values of v' and v'' , may be very large. The change in v , $v' - v''$, however, although not confined to ± 1 as for a harmonic oscillator, tends to be restricted to moderate values. A spectrogram showing parts of three such band

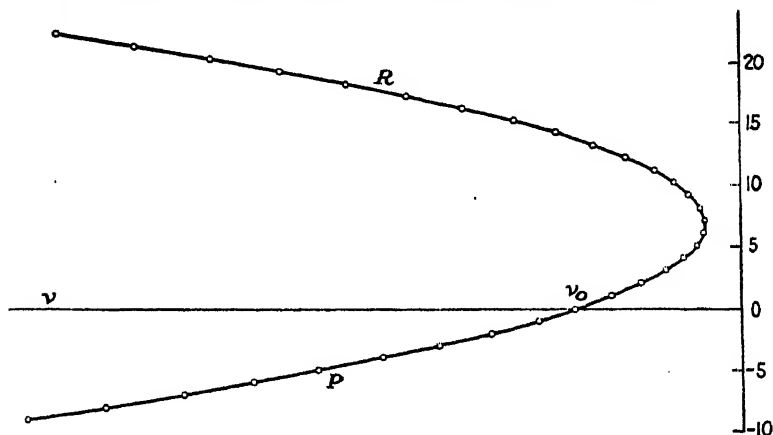
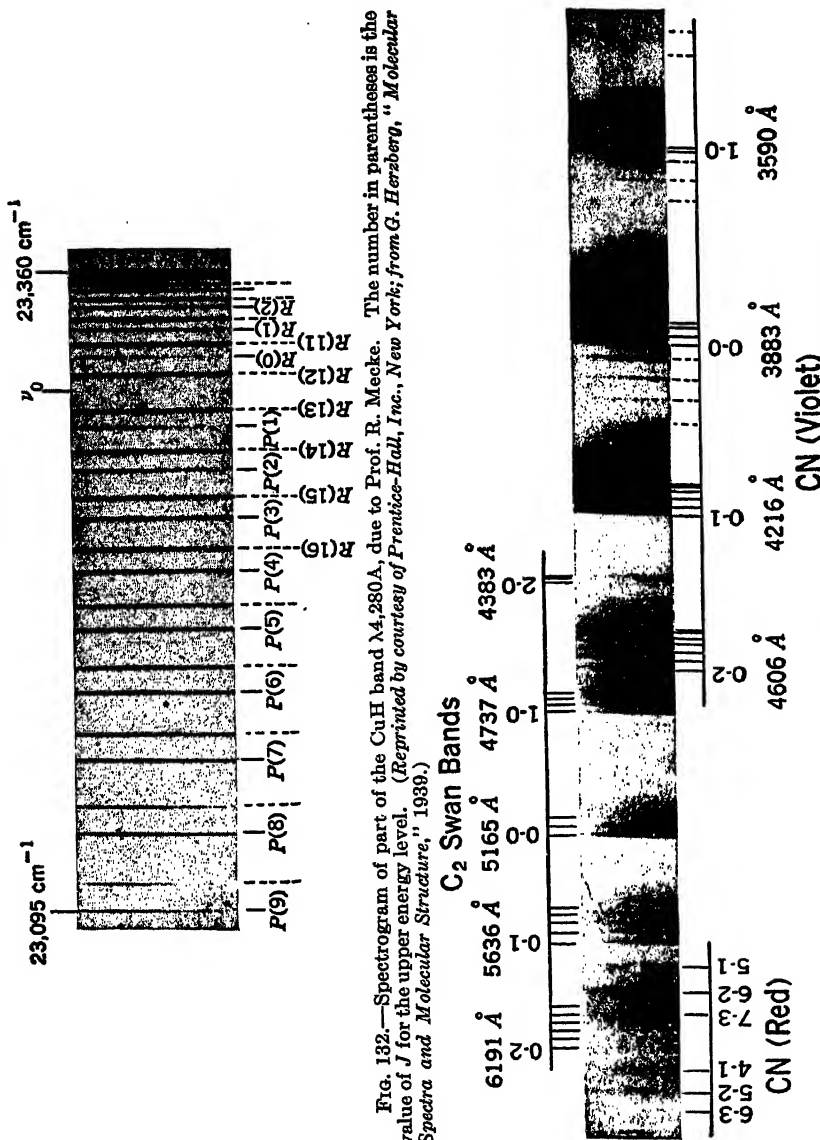


FIG. 131.—Fortrat diagram for the CuH band $\lambda 4,280$ A. The lines are represented by circles, the ordinate representing the larger J value for the transition. $\nu_0 = \nu_{\text{center}}$.

systems is reproduced in Fig. 133. Extremely low dispersion was used in order to include a wide spectral range, hence each *band* appears as a single *line* in the figure; the groups that seem to the eye to stand out are groups of bands having in each case the same value of the difference $v' - v''$. The values of v' , v'' (e.g., $6 - 3$) are shown for a few bands. The Swan bands of C_2 result from a $^3\Pi \rightarrow ^3\Pi$ electronic transition; the violet CN bands, from a $^2\Sigma \rightarrow ^2\Sigma$; the red CN bands, from a $^2\Pi \rightarrow ^2\Sigma$ transition.

The *rotation* spectrum of a molecule like HCl (Sec. 166) can be regarded as the $0 - 0$ (i.e., $v = 0$ to $v = 0$) band, and the *vibration-rotation* band (Sec. 167) as the $0 - 1$ band, of a band system in which the molecule remains in its normal state. In the halogen hydrides, like HCl, this latter is a $^1\Sigma$ state; therefore, $J = K$ and, as already remarked, there is no fine structure. There is also no Q branch, and it is partly for this reason that the band exhibits a gap in the center. Furthermore, there is no band head because the *P* and *R*



branches do not fold back but form a single straight sequence of lines. Thus, bands of this sort are exceptionally simple in form.

Many other interesting features of band spectra could be discussed if space permitted. We shall mention only one more fact. The vibrational and rotational frequencies are affected by the *masses* of the atoms, so that the band spectra of molecules containing *different isotopes* of the same element are slightly different. Extra lines in band spectra have sometimes led to the discovery of rare isotopes, such as the oxygen isotope of atomic weight 18, present in only about 1/630th as great concentration as O^{16} .*

For further information about band spectra the student is referred to the treatises mentioned in a footnote to Sec. 165 above or to the literature.

170. The Raman Effect.—As a final topic in spectroscopy brief mention will be made of the Raman effect.

When light passes through a “transparent” substance, solid, liquid, or gaseous, a certain part of the light is scattered in all directions (the Tyndall effect). The most familiar example is the light from a clear sky. Rayleigh ascribed such effects to scattering by the individual molecules, or by groups of molecules much smaller in linear dimension than the wave length of the light. If the incident light is monochromatic, the scattered light is ordinarily observed to be unchanged in frequency, in accordance with Rayleigh’s theory.

It was shown in 1925 by Kramers and Heisenberg,¹ however, that, according to classical electromagnetic theory, if the scattering electrons in an atom or molecule are in motion, the scattered light should contain other frequencies in addition to that of the incident light. Failure to notice this implication of classical theory had been due to the common assumption that the radiating particles were at rest except as disturbed by radiation. In 1928, independently of the theoretical prediction, Raman and Krishnan discovered the phenomenon experimentally,² in the course of an extensive study of scattering by liquids and solids. The scattering of light with a change in its frequency has been studied extensively since then and is commonly called the Raman effect.

To observe the Raman effect, the incident light should be monochromatic and very intense. The scattered light is then seen to con-

* GIAUQUE and JOHNSTON, *Nature*, vol. 123, p. 318 (1929); MECKE and CHILDS, *Zeits. f. Physik*, vol. 68, p. 362 (1931).

¹ KRAMERS and HEISENBERG, *Zeits. f. Physik*, vol. 31, p. 681 (1925).

² RAMAN and KRISHNAN, *Nature*, vol. 121, p. 501 (1928); RAMAN, *Indian Journ. Phys.*, vol. 2, p. 387 (1928).

tain, besides a line of the same frequency as the incident light, several weak lines of other frequencies. If the frequency ν of the incident light is varied, these other lines move along the frequency axis at the same rate, maintaining constant frequency differences from ν and not changing greatly in intensity. In these respects the Raman lines differ sharply in behavior from fluorescent lines, the frequencies of which are fixed by the scattering substance and which flash out only when the incident frequency falls upon an absorption line of the substance. Thus, in the Raman effect, it is *frequency shifts* in the scattered spectrum that are determined by the nature of the scatterer rather than the frequencies themselves.

The explanation of the Raman effect in terms of quantum theory is very simple. When a photon of frequency ν is "scattered" by an atom or molecule whose quantum state is not altered in the process, the scattered photon has the same frequency as had the incident photon. But it may happen that the atom or molecule is at the same time changed from a state in which its energy is W_1 to a state of different energy W_2 . If this happens, conservation of energy requires that the frequency ν' of the scattered photon be modified so that

$$h\nu' + W_2 = h\nu + W_1;$$

hence,

$$\nu' = \nu + \frac{W_1 - W_2}{h}.$$

In terms of such ideas the Raman effect had been predicted, tentatively, at a much earlier date by Smekal.

In the expression just written for ν' , the term $(W_1 - W_2)/h$ can be interpreted as a frequency ν_{12} that the atom might conceivably emit or absorb in the usual way, in jumping from the first state to the second. Thus for the Raman line we may write

$$\nu' = \nu + \nu_{12}.$$

The difference between the frequency of each Raman line and the frequency of the incident light is thus equal to the frequency of some conceivable emission or absorption line of the scattering atom or molecule.

The *intensity* of a Raman line has nothing to do, however, with the intensity of the emission or absorption line that is thus correlated with it. The selection rules for the two are quite different; transitions that are forbidden in ordinary spectra may occur freely in the Raman effect. This is one reason for its great theoretical interest. According to wave mechanics, a *Raman jump* is possible between two levels *A* and *B* only when there exists at least one third level, *C*, such that ordinary

radiative transitions are allowed between A and C and between B and C. It is almost as if the atom or molecule actually jumped first from A to C and then from C to B. The relative probabilities of the various processes do not correspond to this simple picture, however, nor can it be said that the atom or molecule remains any definite length of time in C, as it would in the production of fluorescence.

By means of these principles the array of lines to be expected in any given case is easily worked out. In the case of the scattering of light by molecules, for example, there may be Raman lines associated with the rotational or vibration-rotation spectra of the molecule, in case it possesses such spectra. In the common type of rotation spectrum, the selection rule is $\Delta J = \pm 1$ (Sec. 166). For the Raman lines associated with this spectrum, therefore, the selection rule will be either $\Delta J = 0$ (i.e., $J \rightarrow J \pm 1$ for a jump from A to C and $J \pm 1 \rightarrow J$ from B to C) or $\Delta J = \pm 2$ (e.g., $J \rightarrow J + 1$ and $J + 1 \rightarrow J + 2$). The case, $\Delta J = 0$, however, involves no change in the molecular energy and hence merely contributes to the ordinary or Rayleigh scattering. Thus, effectively, for Raman lines of purely rotational origin, we must have

$$\Delta J = \pm 2.$$

The incident line, as seen in the spectrum of the scattered light, should thus be accompanied on each side by several lines spaced twice as far apart on the frequency scale as the lines of the rotational spectrum. From Eq. (229), it is easily seen that the innermost lines will be distant from the unmodified line 1.5 times as far as the spacing between the Raman lines themselves [the frequencies of the various lines being proportional to $2(2J - 1)$, with $J \geq 2$, J referring to the upper level].

Besides these *rotational* Raman lines, there should also appear, at a much greater distance and only on the long-wave side, a band corresponding to the ordinary *vibration-rotation* spectrum. This can arise from the following Raman transitions, connected to third levels by allowed transitions as shown, the value of ν being indicated numerically:

$$0, J \rightarrow \begin{cases} 1, J + 1 \\ 1, J - 1 \end{cases} \begin{cases} 1, J + 2 \\ 1, J - 2 \end{cases} \quad 0, J \rightarrow \begin{cases} 0, J + 1 \\ 0, J - 1 \end{cases} \begin{cases} 1, J + 2 \\ 1, J - 2 \end{cases}$$

Thus, for this band, the Raman selection rule is

$$\Delta \nu = 1, \quad \Delta J = 0 \text{ or } \pm 2.$$

Lines for which $\Delta J = 0$, involving almost no change in rotational energy, will coalesce into an intense line in the approximate position of the missing central line of frequency $\nu_{v'v''}$ or ν_{01} in the vibration-rotation band. On each side of this intense line, there should then appear other much fainter lines for $\Delta J = \pm 2$ spaced twice as widely as in the rotational spectrum. In case the vibrational level for $v = 1$ lies so low that a considerable fraction of the molecules are maintained in this state by thermal agitation, a similar but weaker Raman band should be observed on the short-wave side of the exciting line as well, associated with the transition $v = 1 \rightarrow v = 0$.

The Raman spectrum observed in light scattered by gaseous hydrogen halides such as HCl fulfills these predictions. The spectrum of the light scattered by HCl gas when strongly illuminated by light from a mercury arc in glass was studied by Wood and Dieke.¹ They found a line at 4,581.8 Å which they interpreted as the single intense line in a vibration-rotation Raman spectrum excited by the Hg line at 4,047 Å. The frequency difference between these two lines is 2,886.0 cm.⁻¹, which is in excellent agreement with the frequency of the missing central line in the 3.5 μ band, viz., 2,885.4 cm.⁻¹ [Sec. 167(b)]. Much closer to the exciting line and on both sides of it there were also a number of lines which they interpreted as the rotational Raman spectrum excited by the mercury line in question.

TABLE XII.—ROTATIONAL RAMAN SPECTRUM FOR HCl. EXCITED BY λ 4,358 Å

λ	$\tilde{\nu} - \tilde{\nu}_0$	$J \rightarrow J'$	$(\tilde{\nu}_1 + \tilde{\nu}_2)_{\text{inf}}$	λ	$\tilde{\nu} - \tilde{\nu}_0$	$J \rightarrow J'$	$(\tilde{\nu}_1 + \tilde{\nu}_2)_{\text{inf}}$
4,314.7	143.8	4 \rightarrow 2	145.4	0 \rightarrow 2	62.4
				4,377.6	101.1	1 \rightarrow 3	103.9
23.8	183.3	5 \rightarrow 3	186.8				
				85.6	142.7	2 \rightarrow 4	145.4
31.2	232.2	6 \rightarrow 4	228.0				
				94.2	187.5	3 \rightarrow 5	186.8
				4,402.4	229.4	4 \rightarrow 6	228.0
				10.4	271.0	5 \rightarrow 7	269.3
				18.6	312.9	6 \rightarrow 8	310.5
				26.5	353.0	7 \rightarrow 9	351.2

The data obtained by Wood and Dieke for the rotational Raman lines excited by another mercury line, λ 4,358 Å, are reproduced in Table XII, the wave lengths λ of the lines being given in angstroms. Under " $\tilde{\nu} - \tilde{\nu}_0$ " is given the separation in wave numbers of each Raman line from the exciting line; under " $J \rightarrow J'$ " are given the initial and

¹ WOOD and DIEKE, *Phys. Rev.*, vol. 35, p. 1355 (1930).

final values of J ; as $(\bar{\nu}_1 + \bar{\nu}_2)_{\text{inf}}$ is given the sum of the frequencies of the two rotational lines that connect the two states (*e.g.*, for $1 \rightarrow 3$, the frequencies of the rotational lines for $1 \rightarrow 2$ and $2 \rightarrow 3$ are added together). The rotational frequencies are taken from Table XI in Sec. 166; where observed values are missing, calculated values taken from the last column of Table XI are used or others calculated from the same formula there employed. The agreement between $\bar{\nu} - \bar{\nu}_0$ and $(\bar{\nu}_1 + \bar{\nu}_2)_{\text{inf}}$ is sufficiently good, in view of the uncertainty of the infrared data, to constitute strong confirmation of the theory as to the origin of the Raman lines.

Many Raman spectra have been observed. Commonly they are more complicated than the very simple type exhibited by HCl. Even homonuclear molecules, such as O_2 and N_2 , give Raman spectra corresponding to vibration-rotation and rotational spectra, although the latter cannot be observed directly [Sec. 167(*g*)]. The Raman effect thus constitutes a valuable tool in the study of molecular quantum states.

CHAPTER IX

THE QUANTUM THEORY OF SPECIFIC HEATS

In the last few chapters, we have discussed the application of quantum theory to the processes of the emission and absorption of radiation. The next important application of quantum theory, however, after the publication of Bohr's first paper, was made in a different field, in relation to the theory of specific heat. The success of the theory in this field demonstrated beyond doubt the wide range of applicability of the new ideas. In this chapter we shall discuss very briefly the applications of quantum ideas to the theory of specific heats, mainly with the purpose of pointing out how the newer theory succeeded where the old one had largely failed.

For the benefit of students who are not at the moment interested in specific heats, it may be remarked that none of the material in this chapter is necessary for the understanding of later chapters.

171. Variation of Specific Heats of Solids with Temperature.

Over 100 years ago, Dulong and Petit, as a result of the determination of the specific heats of a number of elements, such as iron, lead, silver, gold, concluded that "the product of the atomic weight and the specific heat is the same for all elementary (solid) substances." This law is known as the "law of Dulong and Petit." The product, atomic weight by specific heat, is known as the *atomic heat*.

The law of Dulong and Petit represents at best only a rough approximation, and there are some notable exceptions. The atomic heats in terms of joules are given for 63 elements in crystalline form in the "International Critical Tables." Converted into calories, the atomic heats of 58 of these elements average 6.15 cal. per gram-atom; but they exhibit considerable variation, ranging from 5.38 to 6.93. The atomic heats of the remaining 5 of the 63 elements are in complete disagreement with the law: boron, 3.34; beryllium, 3.85; carbon (diamond), 1.46; hydrogen (solid), 0.57; silicon, 4.95. These 5, with the exception of hydrogen, are light elements having high melting points: boron, 2300; beryllium, 1350; silicon, 1420°C. The specific heats are taken at ordinary temperatures wherever possible.

The exceptions to the law of Dulong and Petit were the subject of extended research, and it was early discovered that at least one reason for the failure of the law in these cases was the fact that the

specific heat varies with temperature. Thus, in 1872, Weber observed that the specific heat of diamond (*i.e.*, carbon) increases threefold between 0 and 200°C. The study of specific heats of substances as a function of temperature has been greatly facilitated in recent years by the availability of liquefied gases for making measurements at low temperatures. It has been found that below a certain temperature characteristic of each substance the specific heat of any solid decreases rapidly with decreasing temperature, ultimately reaching a value almost zero. Figure 134 shows the variation with temperature of the

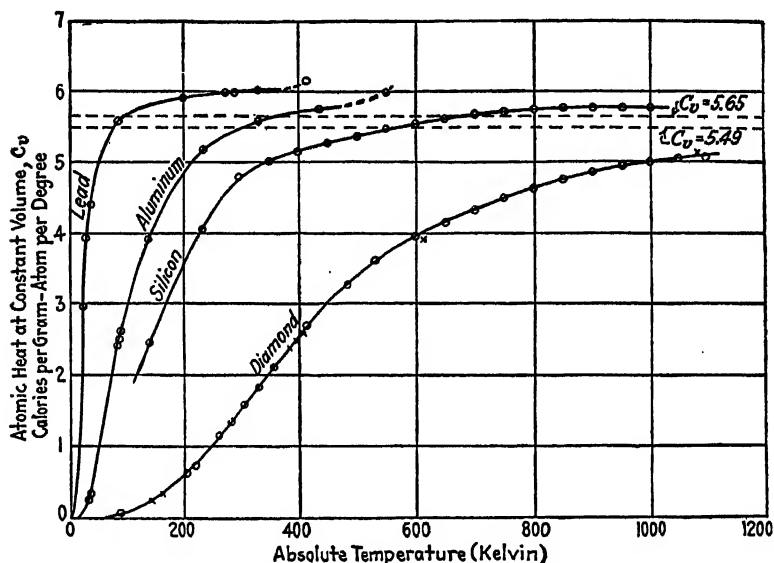


FIG. 134.—Variation of atomic heat at constant volume with temperature.

atomic heats at constant volume of four characteristic substances: diamond, silicon, aluminum, and lead. The atomic heat of lead at ordinary temperatures is observed to approximate the value required by the law of Dulong and Petit, but below a temperature of about 100°K. it drops very rapidly toward zero. The curves for the other three substances have the same general form as that for lead. They differ mainly (1) in the temperature above which they seem to obey the law of Dulong and Petit and (2) in the rate at which the atomic heat decreases below that temperature. In the case of diamond, the curve suggests that a value of about 6 cal. per gram-atom per degree would not be reached until the temperature exceeded 2000°K.

The curves for nearly all substances are similar to those shown in Fig. 134; they lie between the curves for lead on the one side and for

diamond on the other, very commonly even between those for aluminum and for lead. Since most of the measurements on which the law of Dulong and Petit was based were made at *room* temperature, it is obvious why the law is so nearly obeyed.

The curves for different substances are not only similar in appearance but can be brought (almost) into coincidence by suitable change in the temperature scale for each curve. Thus, if the abscissas of the curve for aluminum be multiplied by a factor of 4.6, this curve will nearly coincide with that for diamond. The crosses (X) on the curve for diamond represent the points from the aluminum curve shifted in the way described. This fact indicates that the curves are closely related to each other by some fundamental underlying law. What is this fundamental law which governs the variation of specific heat with temperature and which, at sufficiently high temperatures, must give, approximately, the law of Dulong and Petit? Classical physics failed to find the law. The quantum theory was more successful.

172. Classical Theory of the Specific Heats of Solids.—According to classical conceptions, the atoms of a solid element at the absolute zero of temperature would be at rest in positions of equilibrium under the action of their mutual attractions and repulsions. Let us consider the energy of the solids in this state to be zero. When the temperature is raised, the atoms will be set into vibration about their positions of equilibrium; but as long as the amplitude of vibration is not too large, the restoring force will be proportional to the displacement, and the vibrations will be, therefore, of the simple harmonic type.

Each atom of the solid, like a monatomic gas molecule, has obviously 3 degrees of freedom. According to the principle of the equipartition of energy, discussed in Secs. 84 and 85 of Chap. V, when the solid is at an absolute temperature T , each degree of freedom will have an average amount of energy given by

$$E = kT,$$

k being Boltzmann's constant. That is, for each degree of freedom there is average kinetic energy equal to $\frac{1}{2}kT$, and, because the motion is simple harmonic, there is also potential energy averaging $\frac{1}{2}kT$ as well. The total average energy per atom is thus $3kT$. The total thermal energy of agitation of a gram-atom of the solid, containing N_0 atoms, is, therefore,

$$W_A = 3N_0kT = 3RT, \quad (239)$$

since $N_0k = R$, the gas constant for a gram-atom.

Now the specific heat of a substance is equal to the energy required to raise the temperature of 1 gram 1 degree; hence, the atomic heat is equal to the energy required to raise the temperature of 1 gram-atom 1 degree. If, during this rise in temperature, the substance is not allowed to expand, all of the added energy goes to increase the energy of vibration of the atoms. Calling the value of the atomic heat at constant volume, measured in mechanical units, C_v , it follows that

$$C_v = \frac{dW_A}{dT} = 3R. \quad (240)$$

According to classical theory, then, the atomic heat at constant volume of a monatomic solid should be (1) constant and *independent of temperature* and (2) equal to three times the gas constant R . Since the numerical value of R is 1.9856 cal. per gram-atom per degree, it follows that the numerical value of C_v , on the classical theory of the equipartition of energy, should be

$$C_v = 5.96 \text{ calories per gram-atom per degree.}$$

The values ordinarily given for the specific heat of a solid refer, to be sure, to the specific heat under constant pressure. The difference, however, is at most a few percent.¹ Hence, to a first approximation, the value just found for C_v is the numerical value of atomic heat required by the law of Dulong and Petit.

The value of C_v given in Eq. (240) is obviously very close to the value which the curves in Fig. 134 approach asymptotically as the temperature increases. But Eq. (240) predicts no variation in atomic heat with temperature, as is required by these experimental curves.

173. Einstein's Theory of the Atomic Heats of Solids.—An important advance in the theory of the specific heats of solids was made by Einstein,² who applied the quantum theory to the vibrations which atoms were assumed to execute about their positions of equilibrium. Let it be assumed (1) that the atoms of a monatomic solid vibrate with a frequency ν , the same for all atoms, which depends on the mass of the atom and on the restoring forces brought into play when the atom is displaced, and (2) that the average energy associated with each degree of freedom is that given by the quantum theory for a harmonic oscillator, or, by Eq. (120a) in Sec. 91,

$$\bar{U} = \frac{h\nu}{e^{h\nu/kT} - 1}.$$

¹ See any book on Heat, e.g., J. M. Cork, "Heat," p. 233, 1933; J. K. Roberts, "Heat and Thermodynamics," p. 157, 1933.

² *Ann. d. Physik*, vol. 22, p. 180 (1907).

Since each atom has 3 degrees of freedom, the average energy per atom \bar{E}_a is, then,

$$\bar{E}_a = \frac{3h\nu}{e^{h\nu/kT} - 1},$$

and the total energy W_A of the N_0 atoms in a gram-atom is

$$W_A = N_0 \bar{E}_a = \frac{3N_0 h\nu}{e^{h\nu/kT} - 1}. \quad (241)$$

As before, the atomic heat at constant volume is obtained by differentiating W_A with respect to T . Therefore,

$$C_v = \frac{dW_A}{dT} = 3N_0 h\nu \frac{1}{(e^{h\nu/kT} - 1)^2} e^{h\nu/kT} \frac{h\nu}{kT^2},$$

or, after rearranging and remembering that $N_0 k = R$,

$$C_v = 3R \left[\frac{e^{h\nu/kT}}{(e^{h\nu/kT} - 1)^2} \left(\frac{h\nu}{kT} \right)^2 \right]. \quad (242)$$

This is Einstein's equation, based upon the older quantum theory, for the atomic heat of a solid at constant volume. It differs from Eq. (240), given by classical theory, in that the term $3R$ is multiplied by the quantity within the square bracket, which is a function of temperature. Thus according to the quantum theory C_v is a function of temperature.

Now the experimental curves in Fig. 134 show three main characteristics: (1) they have essentially the same form, as we have seen, for all substances; (2) at low temperatures, the atomic heat approaches zero; (3) at high temperatures, *i.e.*, sufficiently high for each substance, the atomic heat approaches the constant value $3R$.

Einstein's equation, (242), agrees with experiment as regards these three characteristics. (1) According to Eq. (242), the atomic heats of different substances will differ only because of differences in the characteristic frequency ν . At corresponding temperatures, such that the values of the ratio ν/T are the same, the value of C_v will be the same for all substances. (2) If $T \rightarrow 0$, the quantity $h\nu/kT$ becomes infinite, and it may readily be shown that the square bracket in Eq. (242) becomes zero. Einstein's equation, therefore, predicts correctly that the atomic heat should become zero at absolute zero. (3) If kT becomes very large with respect to the product $h\nu$, $h\nu/kT$ approaches zero, and the square bracket approaches unity. At high temperatures, therefore, the atomic heat should approach $3R$ in agreement with experiment and with the law of Dulong and Petit.

For intermediate temperatures, Eq. (242) predicts *qualitatively* the correct variation of atomic heat with temperature, but it yields too low values at low temperatures. This is shown in Fig. 137 below, in which the lower curve is for the atomic heat of aluminum as given by Einstein's formula, whereas the circles are the observed values of C_v used in plotting the curve for aluminum in Fig. 134. The frequency ν in Einstein's formula was chosen so as best to fit the data. In spite of the disagreement, really comparatively small, between theory and experiment, it is remarkable that Einstein's simple theory should predict values of C_v so nearly correct.

174. Characteristic Frequencies.—Einstein's formula for the atomic heat of a solid, given in Eq. (242), contains, in the frequency ν , a single constant characteristic of a given substance. Experimental values of this constant can be obtained from the atomic-heat data by choosing ν so as to obtain the best fit between Einstein's formula and the data. But several independent methods of estimating ν have also been proposed. They lead to values in substantial agreement with those computed from the specific-heat curves.

(a) *From the "Reststrahlen" or "Residual Rays."*—One of the most convincing of these methods is that due to Rubens¹ and collaborators, in which they used the phenomenon of selective reflection for isolating long waves. The reflectivity² R of a substance for radiation of wave length λ depends on both the coefficient of absorption of the substance and its refractive index n at that wave length, the value of R being given for normal incidence by³

$$R = \frac{(1 - n)^2 + n^2\kappa^2}{(1 + n)^2 + n^2\kappa^2};$$

here $\kappa = \lambda\mu/4\pi$, μ being the coefficient of absorption for radiation of wave length λ in the material.⁴ Any substance which exhibits *selective absorption* in any spectral region presumably has resonating mechanisms the natural periods of which coincide with the central part of the absorption band. In those regions, the value of κ and, therefore, of $n\kappa$ may become large enough compared with $(n - 1)$ to raise the value of the reflectivity R considerably above its value for neighboring

¹ RUBENS and NICHOLS, *Ann. d. Physik*, vol. 60, p. 418 (1897); RUBENS and KURLBAUM, *Ann. d. Physik*, vol. 4, p. 649 (1901); RUBENS and HOLLNAGEL, *Phil. Mag.*, vol. 19, p. 761 (1910); RUBENS and VON WARTENBERG, *Königl. Preuss. Akad., Ber.*, p. 169, 1914.

² For definition, see Sec. 74(e).

³ See WOOD, R. W., "Physical Optics."

⁴ Thus μ is not the "absorptivity" as defined in Sec. 74(d). The fraction of the radiant energy that is absorbed in a short distance Δx is $\mu \Delta x$.

spectral regions. The substance then exhibits *selective reflection*. Confirmation of this relation between selective absorption and selective reflection is found in the work of Nichols and of Rubens on quartz. Quartz is transparent, or nearly so, up to about 7.5μ but has two very strong absorption bands near together at about 8.5μ . At 8.5μ , it is found to reflect 80 percent, while at 4μ , where it is quite transparent, it reflects only a few percent.

If a continuous spectrum is reflected from a substance which shows selective reflection at a wave length λ_0 , the reflected beam will be relatively much richer in radiation of that wave length than was the incident beam. After several such reflections, the radiation may be comprised almost entirely of the wave lengths which are selectively reflected, the remainder having been almost completely absorbed. The radiation remaining after several such reflections is called "residual rays" or "reststrahlen."

By this method, Rubens and his collaborators succeeded in isolating the residual rays from a number of substances. A few of their results are shown in Table I, in comparison with values of ν calculated by means of Einstein's formula from the curves for the specific heats of the same substances. The two sets of frequencies are in fair agreement with each other.

(b) *From Compressibilities*.—The frequency of vibration of the atoms, in Einstein's theory, will depend upon the rate at which a restoring force comes into existence as the atoms are displaced from their equilibrium positions. Now the atoms can also be displaced relatively to one other by *compressing* the substance; and, when this is done, the ratio of the stress thus called into existence to the displacement determines the compressibility of the substance. A relation might be expected to exist, therefore, between the characteristic frequency of Einstein's theory and the compressibility. Einstein developed a tentative theory of this relation¹ and used it to calculate values of ν from compressibilities. Two values of ν thus calculated are given in Table I.

(c) *From Melting Points*.—This method of computing frequencies is due to Lindemann.² It is based on the assumption that the melting point T_m of a solid is the temperature at which the amplitude of vibration of the atoms is equal to ("of the same order of magnitude as" would be preferable) the average distance apart d of the atoms. When an atom is executing a simple harmonic motion with amplitude d and frequency ν , we can write for its displacement

¹ *Ann. d. Physik*, vol. 34, pp. 170, 590 (1911).

² *Phys. Zeits.*, vol. 11, p. 609 (1910).

$$s = d \sin 2\pi \nu t.$$

Its energy is entirely kinetic at the end of a swing, when its velocity is

$$\nu = \left(\frac{ds}{dt} \right)_{\max} = 2\pi \nu d (\cos 2\pi \nu t)_{\max} = 2\pi \nu d;$$

hence, its energy of vibration is

$$\frac{1}{2} m \nu^2 = 2\pi^2 m \nu^2 d^2,$$

m being its mass. But by the principle of the equipartition of energy, cited in Sec. 172, combined with Lindemann's assumptions, the energy of the atom also equals $3kT_m$, where T_m is the melting point. Hence,

$$2\pi^2 m \nu^2 d^2 = 3kT_m.$$

Multiply this equation through by N_0 , the number of atoms in a gram-atom, and then substitute $N_0 m = A$, the atomic weight, and $N_0 k = R$, the gas constant for a gram-atom, and we obtain

$$2\pi^2 \nu^2 d^2 A = 3RT_m.$$

We can estimate d as follows. If V_A is the volume occupied by a gram-atom of the substance at temperature T_m , V_A/N_0 is the average volume per atom; supposing this volume to be cubical in form and hence equal, according to Lindemann's assumption, to d^3 , we have

$$d = \left(\frac{V_A}{N_0} \right)^{1/3}.$$

Substituting this value of d in the last equation and solving for ν , we find

$$\nu = C \sqrt{\frac{T_m}{A V_A^{2/3}}}, \quad C = \frac{1}{\pi} N_0^{1/3} \sqrt{\frac{3R}{2}}.$$

TABLE I.—COMPARISON OF CHARACTERISTIC FREQUENCIES

Substance	From resid- ual rays		From compressi- bilities	From melting points	From Einstein's formula
	λ	ν^*	ν^*		
NaCl.....	52.0 μ	5.77	4.6
KCl.....	63.4	4.73	3.7
KBr.....	82.6	3.64	3.2
Aluminum.....	6.6	7.6	6.42
Silicon.....	9.6	11.9
Lead.....	2.2	1.8	1.61
Diamond (extrapolated)...	32.5(?)	30.8

* All frequencies are to be multiplied by 10^{12} sec.⁻¹.

Some values of ν computed from observed melting points by means of this formula are shown in Table I.

The fact that values of ν computed by these diverse methods agree roughly with one another and with those calculated by means of an application of Einstein's formula to the specific heats indicates that there must be a substantial measure of truth in the assumptions underlying all of these theories. Yet it can scarcely be said that any one of these methods of determining the value of ν to be substituted in Einstein's equation leads to a very satisfactory theory of the specific heat.

175. Debye's Theory of Atomic Heats.—An empirical improvement was effected in Einstein's formula by Nernst and Lindemann, who represented the specific heat as the sum of two terms of the Einstein type, each containing a different frequency. This was a step in the right direction. A much more successful theory was developed, however, by Debye, attacking the problem from a different angle.¹

Let us postulate, for the moment, a solid composed of atoms that have no thermal vibrations but are at rest in their respective positions of equilibrium. Let a system of standing waves, say longitudinal, be set up in the solid. Each atom is then vibrating with an amplitude that depends on the position of the atom with respect to the nodes and loops of the wave system. If we superpose more and more standing-wave systems of both the same frequency as the original and of different frequencies, the vibrations of any particular atom become more and more complex, until, finally, we may approach a condition of atomic agitation similar to the temperature vibrations. Conversely, we may think of the temperature vibrations of the atoms of a solid as being equivalent to a vast complex of standing waves of a great range of frequencies.

The set of standing waves thus imagined is similar to the standing waves which we used in Sec. 87 to represent the radiation field inside an enclosure, but it differs in two respects. In the first place, the range of frequencies here certainly does not extend to infinity. For each set of standing waves constitutes a degree of freedom of the solid, and the total number of degrees of freedom is only $3N$, where N is the number of atoms in the solid. There exists, therefore, an upper limit to the frequency of the standing waves. The wave length corresponding to the highest frequency should be of the order of magnitude of the distance between atoms.

¹ *Ann. d. Physik*, vol. 39, p. 789 (1912).

In the second place, two different types of waves are possible in a solid: (1) longitudinal waves, velocity v_L , and (2) transverse waves, velocity v_T . These velocities are determined by the elastic constants of the material and its density.

For the number of degrees of freedom per unit volume dn_L of longitudinal waves in the wave-length range $d\lambda$, we obtained in Sec. 87, Eq. (109a):

$$dn_L = \frac{4\pi d\lambda}{\lambda^4}.$$

The corresponding number dn_T for the transverse waves in the solid will be the same as for transverse electromagnetic waves as given by Eq. (109b) in Sec. 87:

$$dn_T = \frac{8\pi d\lambda}{\lambda^4}.$$

Changing wave length λ to frequency ν by the relations $\lambda = v/\nu$ and $|d\lambda| = (v/\nu^2)d\nu$, these equations become

$$dn_L = 4\pi \frac{\nu^2 d\nu}{v_L^3}, \quad dn_T = 8\pi \frac{\nu^2 d\nu}{v_T^3}.$$

The total number of degrees of freedom dn in unit volume of the solid in the frequency range ν to $\nu + d\nu$ is the sum, $dn_L + dn_T$, since both systems of waves are simultaneously present. Therefore,

$$dn = dn_L + dn_T = 4\pi \left(\frac{1}{v_L^3} + \frac{2}{v_T^3} \right) \nu^2 d\nu \quad (243)$$

Now let us assume that with each of these degrees of freedom of frequency ν there is associated an average amount of energy $\bar{\epsilon}$ equal to

$$\bar{\epsilon} = \frac{h\nu}{e^{h\nu/kT} - 1},$$

by Eq. (120a) in Sec. 91. The total energy dW_A in this frequency range ν to $\nu + d\nu$ in a volume V_A equal to the volume of a gram-atom is then

$$dW_A = V_A \bar{\epsilon} dn = 4\pi V_A \left(\frac{1}{v_L^3} + \frac{2}{v_T^3} \right) \frac{h\nu}{e^{h\nu/kT} - 1} \nu^2 d\nu,$$

and the total energy W_A of the gram-atom, taking into account all frequencies, is

$$W_A = \int dW_A = 4\pi V_A \left(\frac{1}{v_L^3} + \frac{2}{v_T^3} \right) \int_0^\infty \frac{h\nu}{e^{h\nu/kT} - 1} \nu^2 d\nu, \quad (244)$$

where ν_m is the maximum frequency of the standing waves. We may take the lower limit to be 0 because the fundamental frequency of any solid body of experimental dimensions is very small relative to the frequencies of most of the standing waves.

An approximate value of ν_m was obtained by Debye from the condition that the total number of degrees per gram-atom must be $3N_0$, where N_0 is the number of atoms in a gram-atom. The number of the degrees of freedom can also be computed by integrating the expression for dn in Eq. (243) from 0 to ν_m . Hence,

$$3N_0 = 4\pi V_A \left(\frac{1}{v_L^3} + \frac{2}{v_T^3} \right) \int_0^{\nu_m} \nu^2 d\nu = \frac{4\pi V_A}{3} \left(\frac{1}{v_L^3} + \frac{2}{v_T^3} \right) \nu_m^3. \quad (245)$$

Since the velocities are computable from the elastic constants, ν_m can be computed from this equation. By means of Eq. (245), we can also write for W_A , from Eq. (244),

$$W_A = \frac{9N_0}{\nu_m^3} \int_0^{\nu_m} \frac{h\nu}{e^{h\nu/kT} - 1} \nu^2 d\nu. \quad (246)$$

This equation may be compared with Eq. (241) in Einstein's theory.

To bring out the qualitative features of the result, and for purposes of computation, it is convenient to change the variable of integration in Eq. (246) from ν to

$$x = \frac{h\nu}{kT}; \quad \therefore \nu = \frac{kT}{h} x, \quad d\nu = \frac{kT}{h} dx.$$

The upper limit for x is then $h\nu_m/kT$. Since $h\nu$ and kT both have the dimensions of energy, x is dimensionless; hence, $h\nu_m/k$ must have the dimensions of a temperature. It is convenient to call a temperature T_c equal to $h\nu_m/k$ the *characteristic temperature* for the substance under discussion, so that

$$T_c = \frac{h\nu_m}{k}, \quad \therefore \nu_m = \frac{kT_c}{h}. \quad (247a,b)$$

Let us also introduce $R = N_0 k$, the gas constant for a gram-atom. With these substitutions, Eq. (246) takes the form

$$W_A = 9R \frac{T^4}{T_c^3} \int_0^{T_c/T} \frac{x^3}{e^x - 1} dx. \quad (248)$$

To obtain C_v , we differentiate W_A with respect to T , according to the rules for the derivative of a product, remembering that in the second factor, the integral, T appears only in the limit of the integra-

tion. Therefore,

$$C_v = \frac{dW_A}{dT} = 9R \left[4 \frac{T^3}{T_c^3} \int_0^{T_c/T} \frac{x^3}{e^x - 1} dx - \frac{T^4}{T_c^3} \frac{(T_c/T)^3}{e^{T_c/T} - 1} \left(\frac{T_c}{T^2} \right) \right]$$

or

$$C_v = 9R \left[4 \left(\frac{T}{T_c} \right)^3 \int_0^{T_c/T} \frac{x^3}{e^x - 1} dx - \frac{T_c}{T} \frac{1}{e^{T_c/T} - 1} \right]. \quad (249)$$

This is Debye's equation for the specific heat of a solid at constant volume. Like Einstein's equation, (241), it contains, besides the universal constant R , just one additional constant, here T_c , referring to the particular substance under consideration.

176. Experimental Test of Debye's Equation.—1. At *high temperatures*, Debye's equation should yield the classical value of C_v , viz., $3R$. For large values of T , T_c/T becomes small, and we can write, since $e^x = 1 + x + \dots$, dropping all higher terms in the series,

$$e^{T_c/T} - 1 = \frac{T_c}{T}.$$

Similarly, in the integral in (249), x remains small and we can write, replacing $e^x - 1$ by x ,

$$\int_0^{T_c/T} \frac{x^3}{e^x - 1} dx = \int_0^{T_c/T} x^2 dx = \frac{1}{3} \left(\frac{T_c}{T} \right)^3.$$

Hence, by (249), approximately,

$$C_v = 9R \left(\frac{4}{3} - 1 \right) = 3R,$$

as we had anticipated.

2. At *low temperatures* T_c/T is large. Hence, the second term in the square bracket in Eq. (249) becomes negligible and we may replace T_c/T in the upper limit of the integral by infinity. As stated in Sec. 93,

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}.$$

Hence Eq. (249) becomes, approximately,

$$C_v = 9R \left(4 \frac{T^3}{T_c^3} \frac{\pi^4}{15} \right) = \frac{12}{5} \pi^4 R \frac{T^3}{T_c^3} = \frac{234R}{T_c^3} T^3. \quad (250)$$

At temperatures much below its characteristic temperature, therefore, the atomic heat of a solid should be proportional to the *cube* of the absolute temperature. This is confirmed by experiment over a

considerable range of temperatures. Schrödinger¹ quotes extensive data bearing on this point; curves for several of the substances given

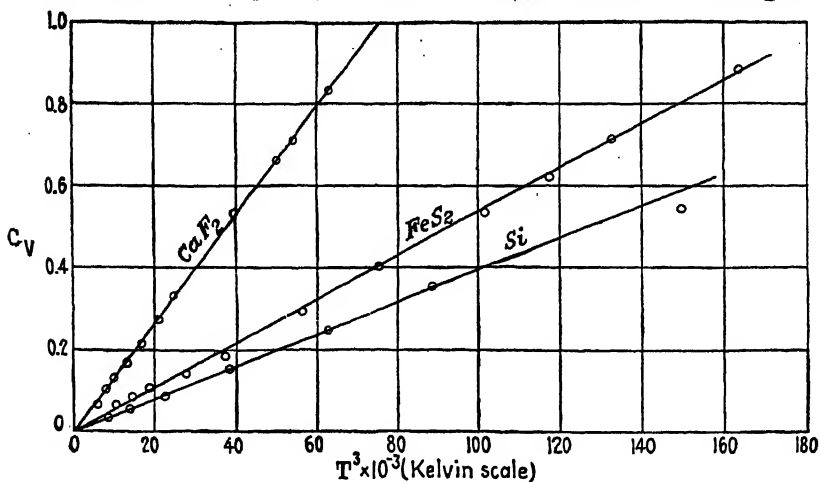


Fig. 135.—At low temperatures, the molecular heat at constant volume is proportional to the cube of the absolute temperature.

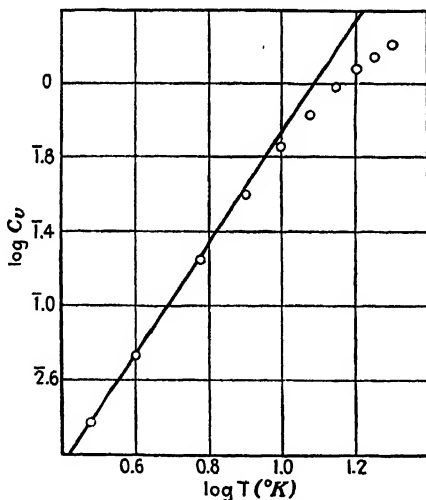


Fig. 136.—Variation of C_v of lead with temperature from 3°K. to 20°K. from measurements of Keesom and Van den Ende.

in his table are shown graphically in Fig. 135, where C_v , denoting the specific heat of a gram-molecule, is plotted against the cube of the absolute temperature for CaF_2 (17 to 40°K.), FeS_2 (22 to 57°K.), and Si (20 to 53°K.). A plot of C_v against $\log T$ for lead, at much

¹ *Phys. Zeits.*, vol. 20, p. 497 (1919).

lower temperatures, is shown in Fig. 136.¹ The predicted proportionality is seen to hold from almost vanishingly small values of C_v well up toward $C_v = 1$. Beyond this point, as theory requires, the curves become increasingly concave toward the axis for T^3 or $\log T$.

3. For *intermediate temperatures*, it is necessary to evaluate Debye's formula by the summation of series. The results of such computation

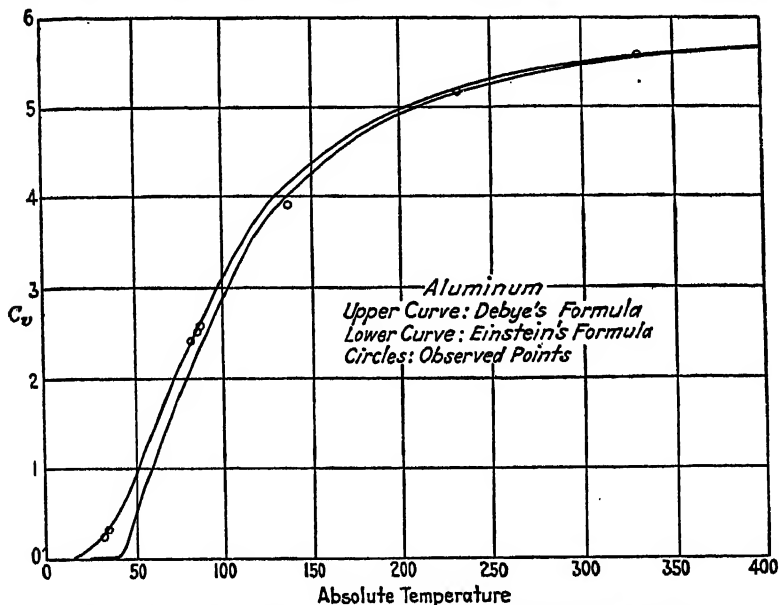


FIG. 137.—Comparison of specific-heat formulas with experiment.

are now available in the form of tables² giving C_v as a function of the quantity T_c/T .

At a temperature $T = T_c$, it is found that

$$C_v = 2.856 R = 5.670 \text{ cal. mole}^{-1} \text{ deg.}^{-1}$$

instead of the classical value, $C_v = 3R$. Thus the critical temperature of a solid, in Debye's theory, might be defined as that temperature at which the atomic heat equals $2.856 R$.

The upper curve of Fig. 137 is the graph of Debye's formula for aluminum, T_c being taken equal to 398° in order to secure the best fit with the data. The agreement between the graph and the observa-

¹ KEESOM and VAN DEN ENDE, *K. Akad. Amsterdam, Proc.*, vol. 33, p. 243 (1930).

² NERNST, "The New Heat Theorem," pp. 246-254, 1924; BEATTIE, *J. Math. Phys.* (Mass. Inst. Tech.), vol. 6, p. 1 (1926).

tions is seen to be excellent. Similar agreement is found for many other substances.¹

Strictly speaking, however, the critical temperature is not an adjustable constant in the Debye theory but is supposed to be given by Eqs. (247a) and (245) in terms of universal constants and the elastic constants of the material. Values of the limiting frequency ν_m for several metals, calculated from the equivalent of Eq. (245) by Allen,² are given in Table II together with values of T_c , labeled "calculated," calculated from these values of ν_m by means of Eq. (247a). For comparison, the values of T_c that give the best fit between Debye's theory and the data³ are added as " T_c , observed." The agreement between the two sets of values of T_c is very good, in view of the approximations made in the theory.

TABLE II.—LIMITING FREQUENCIES ν_m AND CHARACTERISTIC TEMPERATURES T_c .

Element	ν_m	T_c , calculated, °K.	T_c , observed, °K.
Al	8.26×10^{12} sec. ⁻¹	395	398
Fe	9.67	463	453
Cu	6.81	326	315
Ag	4.39	210	215
Pb	1.49	71	88

It is an interesting fact that values of ν_m calculated in this manner from the elastic constants are also of the same order of magnitude as the "characteristic frequencies" obtained in other ways (Sec. 174).

177. Molecular Heats of Mixed Solids.—Excellent as it often is in so far as agreement with experiment is concerned, Debye's theory does not give a complete account of the specific heat of solids. A more complete theory for *crystalline* solids was developed by Born.⁴

A crystal can be regarded as composed of many similar groups or cells of atoms, the groups being spaced uniformly in any given direction. As a result of the mathematical analysis it may be stated that the theory of Debye can be used as a good representation of the motion of the groups as wholes. In addition to this motion, however, there are a finite number of other modes of vibration which can be regarded as modes of vibration of the atoms within each group. For each of

¹ Cf. FOWLER, R. H., "Statistical Mechanics," 2d ed., p. 125, 1936; LEWIS, W. M. C., "Quantum Theory."

² Roy. Soc., Proc., vol. 94, p. 100 (1918).

³ SCHRÖDINGER, Phys. Zeits., vol. 20, p. 450 (1919).

⁴ "Atomtheorie des festen Zustandes," Encyclopädie der mathematischen Wissenschaften, vol. V, part 3, no. 25, 1923; cf. FOWLER, *op. cit.*, p. 119, 1936.

these modes a term in the energy of the Einstein form, as in Eq. (241) above, can be used. The result is a formula combining the expressions given for C_v by the theories of Einstein and Debye, as represented by Eq. (242) and by the equations from which Eq. (249) was obtained:

$$C_v = 9R \frac{d}{dT} \left(\frac{T^4}{T_c^3} \int_0^{T_c/T} \frac{x^3 dx}{e^x - 1} \right) + 3R \sum_{j=1}^n \left(\frac{h\nu_j}{kT} \right)^2 \frac{e^{h\nu_j/kT}}{(e^{h\nu_j/kT} - 1)^2} \quad (251)$$

n being the number of internal modes of vibration of the atoms within a cell. A formula of this type had been suggested empirically by Nernst. At low temperatures the Einstein terms disappear and (251) becomes equivalent to Eq. (249).

By a judicious choice of T_c and of the ν_j 's, this last formula can be made to fit the experimental curve very closely in many cases where the simple Debye theory by itself does not succeed well. It might be anticipated that the frequencies of the residual rays would occur among the ν_j 's. Data confirming this conclusion are shown below; λ_c denotes the wave length corresponding to ν_j as determined by fitting formula (251) to the observed values of C_v , and λ_R is the wave length of the residual rays for the same substance as found by Rubens:¹

	NaCl	KCl	CaF ₂
λ_c	50 μ	61.5 μ	34 μ
λ_R	52	63.4	31.6

For NaCl and KCl, only one ν_j was used in Eq. (251); for CaF₂, two were used, of which one appears to be inactive in optical phenomena.

That characteristic frequencies calculated from *compressibilities* or *melting points* should be of the same order of magnitude as *residual* frequencies, as we have seen them to be (Sec. 174), is understandable if the forces between the atoms of a given cell, which determine the residual frequencies, are of about the same order of magnitude as those between atoms in different cells which also play a role in the determination of compressibilities and melting points. It is not so easy, however, to see just why all of these frequencies should agree roughly, as they actually do, with the limiting frequency ν_m of the Debye theory.

178. The Molecular Heat of Gases: Classical Theory.—The specific heat of a gas of low density presents quite a different problem from that of a solid. The molecules are relatively far apart during

¹ FÜRSTERLING, *Zeits. f. Physik*, vol. 3, p. 9 (1920).

most of the time, so that whatever mutual potential energy they may possess, associated with the forces that they exert on one other, forms only a small part of the total energy. The energy of the gas can thus be regarded as the sum of the energies of the individual molecules.

Each molecule will have, in general, at least three types of motion—translation, rotation, and, for polyatomic gases, vibration of the atoms of the molecule with respect to each other. In addition, we have, of course, electronic vibrations, which, for the present, we shall not consider. Associated with each one of these types of motion we should expect to find a quantity of energy depending on the number of degrees of freedom of each type. The total energy W_M of a gram-molecule at temperature T should, therefore, be

$$W_M = W_t + W_r + W_v, \quad (252)$$

where W_t , W_r , and W_v stand, respectively, for the energy per gram-molecule associated with translation, rotation, and vibration.

Regarding translation, there should be 3 degrees of freedom per molecule, whether the molecule is monatomic or polyatomic. Assuming $\frac{1}{2}kT$ of kinetic energy per degree of freedom, we should have

$$W_t = N_0 \times 3 \times \frac{1}{2}kT = \frac{3}{2}RT. \quad (253)$$

We should also expect to find 3 degrees of freedom of rotation, for the molecule should be capable of rotation about each of three mutually perpendicular axes. Hence, there should be rotational energy per gram-molecule of magnitude $W_r = \frac{3}{2}RT$. Thus, a gas composed of monatomic molecules should have, since obviously $W_v = 0$, total energy per gram-molecule of magnitude

$$W_M = W_t + W_r = \frac{3}{2}RT + \frac{3}{2}RT = 3RT,$$

and its specific heat should be

$$C_v = 3R = 5.96 \text{ calories per mole per degree.}$$

The values of C_v for the monatomic gases helium and argon at -180°C . are observed to be 3.01 cal. per mole;¹ for all of the inert gases at all temperatures, C_v is close to 3. This value is *almost exactly half the value just predicted from classical theory*. To explain this fact, since we know that the molecules of a gas have *translation*, it must be assumed that the *rotation* of such atoms, if rotation exists, does not contribute to the molecular energy. Before the advent of quantum theory, it was supposed that such atoms were spherically symmetrical, so that the forces exerted on them during collisions

¹ "International Critical Tables," vol. V, p. 80.

would act through their centers and hence would never set them into rotation, or would not alter any rotatory motion they might already possess. For a gas composed of such molecules, therefore, we should have

$$W_M = W_t = \frac{3}{2} RT = 2.98T,$$

$$C_v = 2.98 \text{ calories per mole per degree,}$$

in good agreement with the experimental values.

A *diatomic* molecule might then be supposed to consist of 2 atoms *bound rigidly* together, forming a "dumbbell" molecule. If the line through their centers is assumed to constitute an axis of symmetry, any rotational motion about this axis that might exist would never change as a result of collisions and therefore would contribute nothing to the specific heat. As was stated in Sec. 84, however, there would be 2 degrees of freedom corresponding to rotation about each of two mutually perpendicular axes passing through the center of gravity of the molecule at right angles to the axis of the molecule. These 2 degrees of freedom should contribute kinetic energy of rotation of amount $W_r = 2 \times \frac{1}{2} kT = kT$. Thus, for a gram-atom of a gas composed of rigid symmetrical dumbbell molecules, we should have

$$W_M = W_t + W_r = \frac{3}{2} RT + RT = \frac{5}{2} RT$$

and

$$C_v = \frac{5}{2} R = 4.96 \text{ calories per mole per degree.}$$

The measured molecular heats for six diatomic gases are given below:¹

Gas.....	H ₂	N ₂	O ₂	CO	HCl	Cl ₂
C_p	4.84	4.94	4.98	4.94	5.02	6.02

Five of these values of C_p agree well with the value just deduced theoretically for rigid dumbbell molecules. The experimental value for chlorine, however, is much too high. Some explanation for this discrepancy must be found.

An obvious way in which higher values of C_v might arise is by the occurrence of energy of vibration of the atoms relative to each other. As we have seen in Secs. 84 and 85, the average energy for a vibrational degree of freedom is $2E_1 = 2 \times \frac{1}{2} kT = kT$. Thus, for a gram-atom of *vibrating* diatomic molecules, we have $W_v = RT$; and, for its total thermal energy and specific heat, we have

¹ C_v is calculated as $C_v = C_p/\gamma$ from data in the "International Critical Tables," *loc. cit.*

$$W_M = W_t + W_r + W_v = (\frac{3}{2} + 1 + 1)RT = \frac{5}{2}RT,$$

$$C_v = \frac{5}{2}R = 6.95 \text{ calories per mole per degree.}$$

This, however, is *too large* to agree with the experimental value for chlorine! The specific heat of chlorine thus presents an outstanding difficulty for the classical theory.

A *rigid triatomic* molecule should certainly have 3 degrees of freedom in respect of rotation, if we assume that all 3 atoms do not lie in a straight line. Hence, for a gram-atom of such molecules,

$$W_M = W_t + W_r = (\frac{3}{2} + \frac{3}{2})RT = 3RT,$$

$$C_v = 3R = 5.96 \text{ calories per mole per degree.}$$

Actually, the molecular heat of water vapor is 5.96; that for CO_2 at 20°C . is about 6.2. The specific heat of gases the molecules of which contain more than 2 atoms, however, is usually well above $3R$. The excess could easily be accounted for by supposing that in such molecules the atoms are capable of vibration. But then it is hard to understand why atomic vibration should not occur in H_2O and CO , as well, the energy of which is fully accounted for by translation and rotation alone.

Thus classical theory, by means of special assumptions as to molecular structure, could account for the specific heats of many gases but not for all. Another respect in which the theory failed was in regard to the *variation of specific heat with temperature*. There is no reason, according to classical theory, to anticipate any variation at all. Yet experiment showed that the specific heats of apparently all polyatomic gases exhibit marked variation with temperature when observations are extended over a sufficiently wide range of temperature.

179. Quantum Theory of the Specific Heat of Gases.—The key to the solution of the difficulties just discussed, as to so many others, was eventually furnished by the quantum theory. It turned out that the *translational* energy of the molecules can be treated adequately by classical methods, but this is not true of their *internal energy*, in which is included the energies of rotation and of atomic vibration.

The quantum theory of the specific heat of gases can be developed in either of two slightly different forms. In the first form, the internal energy belonging to each quantum state is treated as a unit, and no approximations are made. In the second form, the molecular energy is divided approximately into parts corresponding to the classical subdivision, as in the theory of band spectra.

(a) (*Relatively*) *Exact Theory of Molecular Specific Heats.*—Let us assume, as in the explanation of atomic and molecular spectra, that

each molecule, in addition to its motion of translation, is capable of existing in any one of a large number of internal quantum states. Let its energy when in state number j be ϵ_j . Then, when a gram-molecule of gas is in thermal equilibrium at temperature T , the N_0 molecules composing it will be distributed among their internal quantum states according to Boltzmann's formula. According to Eq. (148a) in Sec. 110, the number of molecules in state number j will be

$$N_j = C e^{-\epsilon_j/kT},$$

C being a constant. Or, if states having the same energy are grouped together, the number in composite state number τ will be, as in Eq. (148c),

$$N_\tau = C w_\tau e^{-\epsilon_\tau/kT}, \quad (254a)$$

w_τ being the statistical weight of that state and ϵ_τ the energy of a molecule when in that state. In the last equation, the value of the constant C is fixed by the condition that the total number of the molecules is N_0 . Hence, \sum_τ denoting a sum over all quantum states,

$$\sum_\tau N_\tau = C \sum_\tau w_\tau e^{-\epsilon_\tau/kT} = N_0.$$

Thus, we can also write, in place of (254a),

$$N_\tau = \frac{N_0 w_\tau e^{-\epsilon_\tau/kT}}{\sum_\kappa w_\kappa e^{-\epsilon_\kappa/kT}}, \quad (254b)$$

the index of summation being changed here from τ to κ in the sum in order to avoid confusion with τ elsewhere.

An expression for the internal energy of the molecules can now be written down. The energy of N_τ molecules in state number τ is $N_\tau \epsilon_\tau$; hence, the internal energy W_i of the molecules in a gram-molecule of gas is

$$W_i = \sum_\tau N_\tau \epsilon_\tau = \frac{N_0 \sum_\tau \epsilon_\tau w_\tau e^{-\epsilon_\tau/kT}}{\sum_\tau w_\tau e^{-\epsilon_\tau/kT}},$$

by (254b). In order to find the total energy of the gas, we have to add to W_i the classical value of the translational energy, which, as given in Eq. (253), is $W_t = \frac{3}{2}RT$. The specific heat is then

$$C_v = d(W_t + W_i)/dT$$

or

$$C_v = \frac{3}{2}R + N_0 \frac{d}{dT} \frac{\sum_r \epsilon_r w_r e^{-\epsilon_r/kT}}{\sum_r w_r e^{-\epsilon_r/kT}}. \quad (255)$$

Now the molecular energies ϵ_r or ϵ_j that occur here are the energies of the same molecular states that are involved in the production of *molecular spectra*. It should be possible, therefore, to take values of ϵ_j derived from a study of band spectra and to calculate C_v from them, by means of Eq. (255). Thus a *specific heat would be calculated from spectroscopic data*. This has actually been done in a number of cases, and with complete success. Such calculations constitute a striking confirmation of the correctness of the general basis underlying the quantum theory of molecular structure.

(b) *Approximate Theory of the Molecular Heats of Gases*.—In order to arrive at a general explanation of the observed features of gaseous specific heats, it is necessary now to introduce the fact, that, as explained in Secs. 167 and 168, the internal energy of a molecule can commonly be divided approximately into separate parts, in analogy with the classical picture of molecular energy as described above. We can write, approximately,

$$\epsilon_r = \epsilon_j + \epsilon_v + \epsilon_n, \quad (256)$$

where ϵ_r denotes *energy of rotation*; ϵ_v , *energy of atomic vibration*; and ϵ_n , *electronic energy* associated with the electrons in the molecule.

Of these three parts, the electronic energy remains constant as long as the molecules are in thermal equilibrium at any temperature that can be realized in the laboratory; molecular collisions due to thermal agitation are not sufficiently violent to excite in appreciable number electronic states above the one of lowest energy. Hence, the existence of ϵ_n has no effect upon the specific heat. The other two parts of the energy, summed for all molecules, constitute, respectively, rotational energy of magnitude $W_r = \sum \epsilon_r$ and vibrational energy of magnitude $W_v = \sum \epsilon_v$. Thus, that portion of the energy of a gram-molecule of gas which varies with temperature can be written as the sum of only three parts, as in Eq. (252) above, or as

$$W_M = W_t + W_r + W_v,$$

W_t being as before the kinetic energy of the translatory motion. It is convenient to divide the molecular heat as well into three corresponding parts:

$$C_v = C_{vt} + C_{vr} + C_{v\tau}; \quad C_{vt} = \frac{dW_t}{dT}, \quad C_{vr} = \frac{dW_r}{dT}, \quad C_{v\tau} = \frac{dW_\tau}{dT}.$$

These three parts of W_M and of C_v may then be discussed separately.

The *translational* energy and heat are, as in classical theory,

$$W_t = \frac{3}{2}RT, \quad C_{vt} = \frac{3}{2}R. \quad (257a,b)$$

A monatomic gas has no other kind of energy; hence for it $C_v = \frac{3}{2}R$, in agreement with experiment. No gas at low density can have (or has) a smaller specific heat than this.

The *rotational* energy, also, can be handled independently. To do this we need first to study the distribution law of the molecules from the standpoint of the rotational motion considered separately.

The three parts into which the energy is divided in Eq. (256) correspond, so to speak, to three sets of "partial" molecular quantum states; there are rotational, vibrational, and electronic quantum states, and a true quantum state for the molecule is a combination of one of each of these three. We may ignore the electronic state, however, since that never changes. Let the statistical weights (or number of degenerate fundamental quantum states) for the (composite) rotational and vibrational states be w_J and w_v , respectively. Then $w_\tau = w_J w_v$. Thus, the number of molecules in molecular quantum state number τ is, from (254a) and (256),

$$N_\tau = w_J w_v C' e^{-(\epsilon_J + \epsilon_v)/kT},$$

the factor $e^{-\epsilon_v/kT}$ being included here in the constant C' . The total number of molecules in rotational state number J , symbol N_J , can then be found by adding up molecules in all possible vibrational states, which gives

$$N_J = \left(C' \sum_v w_v e^{-\epsilon_v/kT} \right) w_J e^{-\epsilon_J/kT}.$$

Here the quantity in parentheses represents a new constant of proportionality. Its value can be found from the condition that

$$\sum_J N_J = N_0,$$

which gives

$$N_0 = \left(C' \sum_v w_v e^{-\epsilon_v/kT} \right) \sum_J w_J e^{-\epsilon_J/kT}.$$

Solving this last equation for the quantity in parentheses and substituting in the preceding equation, we obtain

$$N_J = \frac{w_J N_0 e^{-\epsilon_J/kT}}{\sum_J w_J e^{-\epsilon_J/kT}} \quad (258)$$

for the number of molecules in rotational state number J . This is exactly the same distribution law that would hold if the molecules differed in no other respect than in their rotational states.

The rotational energy of a gram-molecule is then $W_r = \sum_J \epsilon_J N_J$; and the rotational part of the molecular heat is

$$C_{vr} = \frac{dW_r}{dT} = N_0 \frac{d}{dT} \frac{\sum_J \epsilon_J w_J e^{-\epsilon_J/kT}}{\sum_J w_J e^{-\epsilon_J/kT}} \quad (259)$$

This formula for C_{vr} predicts an interesting variation of C_{vr} with temperature. At very *low* temperatures C_{vr} should become vanishingly small. For, if T is small enough, the Boltzmann factor $e^{-\epsilon_J/kT}$ for the state of lowest ϵ_J will be immensely larger than it is for any higher state, even for the next higher. Both sums then reduce effectively to their respective first terms, in which $J = 0$. Thus, the last fraction in (259) becomes

$$\epsilon_0 w_0 e^{-\epsilon_0/kT} / w_0 e^{-\epsilon_0/kT} = \epsilon_0$$

and is independent of T , so that $C_{vr} = 0$.

At *high* enough temperatures, on the other hand, it can be shown that C_{vr} approximates to the classical value. For a molecule of the rigid diatomic or dumbbell type, the proof is so simple that we may give it here. For such a molecule, as was stated in Eq. (229) in Sec. 166,

$$\epsilon_J = J(J+1)B, \quad B = \frac{h^2}{8\pi^2 I} \quad (260a,b)$$

I being the moment of inertia of the molecule about an axis perpendicular to its axis of symmetry. In this case $w_J = 2J+1$ (i.e., the number of different values of the quantum number M). If T is large enough, many terms will contribute appreciably to the sum, the variation from one term to the next being quite gradual except for the first few terms; hence, the sums can be replaced with sufficient approximation by integrals. Thus

$$\sum_J w_J e^{-\epsilon_J/kT} = \sum_{J=0}^{\infty} (2J+1) e^{-B J(J+1)/kT} = \int_0^{\infty} (2x+1) e^{-Bx(x+1)/kT} dx$$

approximately. The equivalence of the integral to the sum is evident from Fig. 138, in which the shaded area composed of rectangles represents the sum, whereas the area under the curve represents the

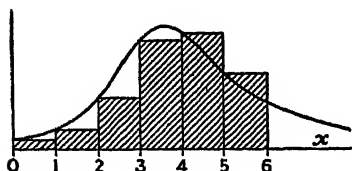


FIG. 138.

integral. (The figure is merely qualitative, and is drawn for such a low temperature that the approximation would actually be poor.) In a similar way we find, integrating once by parts, that

$$\begin{aligned} \sum_J \epsilon_J W_J e^{-\epsilon_J/kT} &= B \sum_{J=0}^{\infty} J(J+1)(2J+1) e^{-BJ(J+1)/kT} = \\ B \int_0^{\infty} (x^2+x)(2x+1) e^{-B(x^2+x)/kT} dx &= -kT(x^2+x) e^{-B(x^2+x)/kT} \Big|_{x=0}^{x=\infty} \\ &+ kT \int_0^{\infty} (2x+1) e^{-B(x^2+x)/kT} dx = kT \int_0^{\infty} (2x+1) e^{-Bx(x+1)/kT} dx. \end{aligned}$$

Upon substituting these values of the sums in (259) we find that the integral cancels out and

$$C_{vr} = N_0 \frac{d}{dT}(kT') = N_0 k = R.$$

This is the classical value as found above (total $C_v = C_{vt} + C_{vr} = \frac{5}{2}R$, $C_{vt} = \frac{3}{2}R$, $\therefore C_{vr} = R$).

At intermediate temperatures C_{vr} has a value between 0 and R which can be calculated by evaluating the sums in Eq. (259).

The *vibrational* energy can be treated in a similar way, provided an expression can be found, or is assumed, for W_v . At low temperatures, all molecules will remain in their lowest vibrational states, and $C_{vv} = 0$. Since the spacing of the vibrational levels is commonly many times that of the rotational levels, C_{vv} will begin to increase from 0 only at a much higher temperature than C_{vr} . Eventually, however, C_{vv} likewise approaches the classical value, which, for a vibrating diatomic molecule, we have found to be equal to R .

Thus the quantum theory predicts the following qualitative mode of variation of C_v with temperature for a polyatomic gas of low density. At very low temperatures, $C_v = \frac{3}{2}R$ for all gases. As T increases, C_v rises (as molecular rotation develops) toward the value of C_v as deduced from classical theory with the inclusion of rotational motion. Thereafter, C_v will remain practically constant again, until molecular vibrations begin to set in, whereupon a further rise in C_v will occur. At very much higher temperatures, a third rise may be expected for any gas, even the monatomic ones, as electronic excitation begins to occur; but temperatures of this order are not at present available in the laboratory.

180. Comparison of the Theory with Observed Specific Heats.—So far as is known, the specific heats of gases vary with temperature in the manner just deduced from theory. The most extensive data exist for hydrogen¹; the observed curve for this gas is shown in Fig. 139,

¹ Cf. "International Critical Tables," vol. V, pp. 79, 84.

the absolute temperature being shown on a logarithmic scale. The shape of the curve agrees with the theoretical prediction.

From *b* to *c* (30 to 60°K.), only the translational motion of the molecules varies as the temperature rises; if there is rotational energy present, it is constant. Presumably this part of the curve continues in a horizontal direction, as indicated by *b* . . . *a*, to 0°K. In this region, hydrogen behaves like a monatomic gas.

At *c* (about 60°), a rise begins, owing to an increasing amount of rotational energy. At *d* (270°K.), the rotational energy has attained

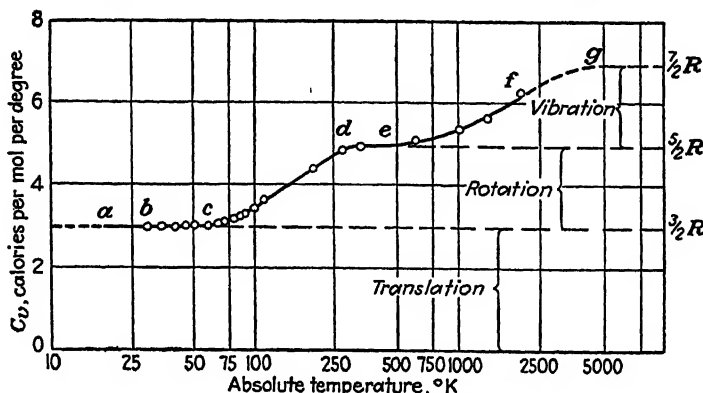


FIG. 139.—Specific heat at constant volume of ordinary hydrogen plotted against temperature on a logarithmic scale.

its full classical value equal to R ; and, from here to *e* (500°K.), C_v remains constant again, at very nearly the classical dumbbell value, $C_v = \frac{5}{2}R$. Beyond *e*, vibration begins, an appreciable fraction of the molecules beginning to exist in their next higher vibrational state.

The approximate constancy of C_v for hydrogen at "ordinary" temperatures is thus seen to be of the nature of an accident, resulting from the fact that the earth's atmosphere has a temperature of some 290°K.

The complete theory of the hydrogen molecule is, however, more complicated than as just described. There are *two forms* of the molecule; these represent alternative quantum states for the pair of protons in it, in analogy with the singlet and triplet states for 2 electrons in an atom (Sec. 144, 150). Hydrogen composed entirely of molecules of one of these two kinds is called parahydrogen, that composed of the other, orthohydrogen. Only 25 percent of ordinary hydrogen is parahydrogen; but, if ordinary hydrogen is adsorbed on charcoal cooled to 20°K. (by means of liquid hydrogen), it becomes converted in the course of a few hours into practically pure parahydrogen.

In most respects, the two forms of hydrogen possess similar physical and chemical properties, but they differ markedly in their band spectra and in the mode of variation of their specific heats with temperature. These differences arise from the fact that the rotational states of parahydrogen are restricted to even values of J , those of orthohydrogen to odd values. As a consequence, below 50°K . the molecules of orthohydrogen possess rotational energy which does not vary with temperature, J being effectively limited to the fixed value $J = 1$, whereas the molecules of parahydrogen are not rotating ($J = 0$, effectively). We have not space here to pursue this interesting subject further.¹

As a concluding illustration of the application of quantum theory to the specific heats of gases, we may return to the problem of chlorine, which was left unsolved in Sec. 178. Whereas the second vibrational state ($v = 1$) for gases such as N_2 , H_2 lies far above the first ($v = 0$), for chlorine it lies only 560 cm.^{-1} above the first, as may be inferred from a study of the electronic band system (Secs. 168 and 169) that extends in the absorption spectrum of chlorine from $4,800$ to $5,800\text{ \AA}$.^{*} The difference in energy between these two states is thus

$$\epsilon_1 - \epsilon_0 = h\nu = 560 \times 3 \times 10^{10} \times 6.61 \times 10^{-27} = 1.11 \times 10^{-13}\text{ erg.}$$

Since $k = 1.38 \times 10^{-16}$, at $T = 288^\circ\text{K}$. the ratio of the Boltzmann factors for the two states is $e^{-(\epsilon_1 - \epsilon_0)/kT} = e^{-2.80} = 0.061$. Thus, 0.061 times as many molecules will be in the second vibrational state as in the first. For a rough estimate of the effect upon the specific heat, we may suppose that the fraction $e^{-(\epsilon_1 - \epsilon_0)/kT}$ of all of the molecules are in the second state. Then the excess vibrational energy due to this cause, in a gram-molecule of gas containing N_0 molecules, will be

$$N_0 (\epsilon_1 - \epsilon_0) e^{-(\epsilon_1 - \epsilon_0)/kT};$$

and the derivative of this expression with respect to T gives for the contribution of the excess energy to the molecular heat

$$N_0 \frac{(\epsilon_1 - \epsilon_0)^2}{kT^2} e^{-(\epsilon_1 - \epsilon_0)/kT} = R \left(\frac{\epsilon_1 - \epsilon_0}{kT} \right)^2 e^{-(\epsilon_1 - \epsilon_0)/kT} = 0.47 R,$$

where $R = N_0 k$. Adding the usual $\frac{5}{2}R$ for a diatomic molecule, we have then $C_v = 2.97R = 2.97 \times 1.98 = 5.88$ calories per mole per degree. This agrees fairly well with the observed value of 6.02 as given in Sec. 178. A more exact theoretical calculation, by the

¹ Cf. KENNARD, E. H., "Kinetic Theory of Gases," p. 262, 1938; FARKAS, "Orthohydrogen, Parahydrogen and Heavy Hydrogen," 1935.

^{*} KUHN, *Zeits. f. Physik*, vol. 39, p. 77 (1926).

method described in Sec. 179(a) above,¹ gives $C_v = 6.06$. Thus quantum theory succeeds where classical theory failed.

When a similar calculation is made for HCl, whose vibration-rotation band at 3.4μ or $2,886 \text{ cm.}^{-1}$ [Sec. 167(b)] indicates that $\epsilon_1 - \epsilon_0 = 2,886 \text{ cm.}^{-1} = 5.72 \times 10^{-13} \text{ erg}$, the ratio of the Boltzmann factors is found to be only $e^{-(\epsilon_1 - \epsilon_0)/kT} = 5 \times 10^{-7}$. Thus, molecular vibration can contribute nothing appreciable to the specific heat of HCl, in agreement with the observed fact that for it $C_v = \frac{5}{2}R$ very nearly. For O_2 , $\epsilon_1 - \epsilon_0 = 1,556 \text{ cm.}^{-1}$, and vibration may contribute about $0.026R$ to C_v .

¹ TRAUTZ and ADER, *Zeits. f. Physik*, vol. 89, p. 15 (1934).

CHAPTER X

X-RAYS

There is probably no subject in all science which illustrates better than X-rays the importance to the entire world of research in pure science. Within 3 months after Roentgen's fortuitous discovery, X-rays were being put to practical use in a hospital in Vienna in connection with surgical operations. The use of this new aid to surgery soon spread rapidly. Since Roentgen's time, X-rays have completely revolutionized certain phases of medical practice. Had Roentgen deliberately set about to discover some means of assisting surgeons in reducing fractures, *it is almost certain that he would never have been working with the evacuated tubes, induction coils, and the like, which led to his famous discovery.*

In other fields of applied science, both biological and physical, uses have been found for X-rays, which approximate in importance their use in surgery. One may mention, for example, the study of the crystal structure of materials; "industrial diagnosis," such as the search for defects in the materials of engineering; the detection of artificial gems or of overripe fruit; the study of old paintings; the study of genetics by the biologist; the use of X-rays by the doctor in radiography and in radiotherapy; and many other uses.

But transcending these uses in applied science are the applications of X-rays made to such problems as the atomic and the molecular structure of matter and the mechanism of the interaction of radiation with matter. X-rays provide us with a kind of supermicroscope, by means of which we can "see" not only atoms and their arrangement in crystals but also even the interior of the atom itself. Roentgen's discovery must be ranked with the most important scientific discoveries of all time.

In this chapter, we shall give a brief account of the development and the present status of X-rays, with particular reference to their application to some of the fundamental problems of physics.

EARLY, MOSTLY QUALITATIVE DEVELOPMENTS IN X-RAYS (1895-1912)

181. Roentgen's Discovery.—In the autumn of 1895, Wilhelm Konrad Roentgen, professor of physics at Würzburg, was studying that fascinating phenomenon, the discharge of electricity through

rarefied gases. A large induction coil was connected to a rather highly evacuated tube (Fig. 140), the cathode *C* being at one end and the anode *A* at the side. The tube was covered "with a somewhat closely fitting mantle of thin black cardboard."¹ With the apparatus in a completely darkened room, he made the accidental observation that "a paper screen washed with barium-platino-cyanide lights up brilliantly and fluoresces equally well whether the treated side or the other be turned toward the discharge tube."¹ The fluorescence was observable 2 meters away from the apparatus. Roentgen soon convinced himself that the agency which caused the fluorescence originated

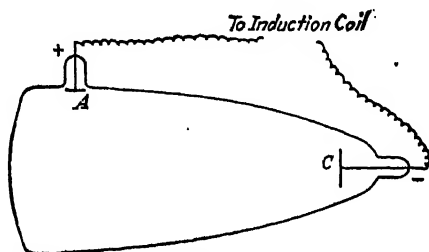


FIG. 140.—Diagram of the tube with which Roentgen discovered X-rays.

at that point in the discharge tube where the glass walls were struck by the cathode stream in the tube.

Realizing the importance of his discovery, Roentgen at once proceeded to study the properties of these new rays—the unknown nature of which he indicated by calling them "X-rays." In his first communications he recorded, among others, the following observations:

1. All substances are more or less transparent to X-rays. For example, wood 2 to 3 cm. thick is very transparent. Aluminum 15 mm. thick "weakens the effect considerably, though it does not entirely destroy the fluorescence." Lead glass is quite opaque, but other glass of the same thickness is much more transparent. "If the hand is held between the discharge tube and the screen the dark shadow of the bones is visible within the slightly dark shadow of the hand."

2. Many other substances besides barium-platino-cyanide fluoresce—calcium compounds, uranium glass, rock salt, etc.

3. Photographic plates and films "show themselves susceptible to X-rays." Hence, photography provides a valuable method of studying the effects of X-rays.

¹ Quotations from papers by Roentgen, *Electrician*, vol. 36, pp. 415, 850 (1896).

4. X-rays are neither reflected nor refracted. Hence, "X-rays cannot be concentrated by lenses." (Subsequent research, however, has demonstrated that both reflection and refraction can be observed under special conditions).

5. Unlike cathode rays, X-rays are not deflected by a magnetic field. They travel in straight lines, as Roentgen showed by means of "pinhole" photographs.

6. X-rays discharge electrified bodies,¹ whether the electrification is positive or negative.

7. X-rays are generated when the cathode rays of the discharge tube strike any solid body. A heavier element, such as platinum, however, is much more efficient as a generator of X-rays than is a lighter element, such as aluminum.

It is a stirring tribute to Roentgen's masterly thoroughness that most of the basic properties of X-rays were described in the paper in which the discovery was first announced.

Roentgen's discovery excited intense interest throughout the entire scientific world. His experiments were repeated and extended in very many laboratories in both America and Europe.² This early work is beautifully illustrative of the qualitative phase of development of a typical field of physics; it is with special interest, therefore, that we review in the next four sections a few of the outstanding experiments and theories of the period from 1895 to about 1912.

182. Production and Measurement of X-Rays. (a) *Tubes.*—Early tubes for the production of X-rays very soon became more or less standardized along the lines suggested by Roentgen in his second paper. Figure 141(a) shows such a tube. A residual gas pressure of the order of 10^{-3} mm. Hg provides, when voltage is applied, a few electrons and positive ions. These positive ions, bombarding the cathode *C*, release electrons—or "cathode rays," as they were called then—and the electrons, hurled against the anode *A*, give rise to X-rays. A curved cathode converges the electrons into a focal spot on *A* of desired shape and size. In this type of tube, known as the "gas" tube, the anode current, applied voltage, and gas pressure are more or less interdependent, and it is essential that the gas pressure be maintained at the desired value. Various ingenious devices were

¹ This property of X-rays was independently discovered by Sir. J. J. Thomson [see letter to *Electrician* (Feb. 4, 1896)]. Thomson pointed out that this phenomenon provides a method of studying X-rays much more delicate and expeditious than the photographic plate or the fluorescent screen and, further, that it yields quantitative measurements.

² The *Beiblätter zu den Ann. d. Physik* for 1896 contain 400 titles on X-rays.

introduced for accomplishing this. These gas tubes were practically the only source of X-rays until the introduction of the Coolidge tube¹ in 1913 and are still extensively used.

The Coolidge type of X-ray tube, illustrated in Figure 141(b), is evacuated to the highest attainable vacuum and incorporates as the essential part of the cathode a hot spiral filament *F* of tungsten of which the temperature, or electron emission, can be separately controlled. This distinguishing feature of the Coolidge tube, *viz.*, control of the tube current independently of the applied voltage, has greatly facilitated certain types of X-ray research.

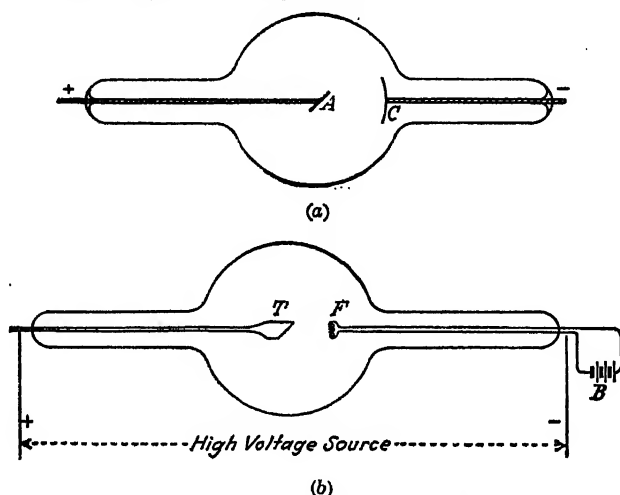


FIG. 141.—(a) An early form of X-ray tube; (b) the Coolidge X-ray tube.

(b) *Measurement of X-ray Intensity.*—Methods for detecting X-rays and for studying qualitatively their relative intensities were based upon the fluorescent and photographic effect of X-rays as reported by Roentgen. The photographic method was especially convenient in medical radiographic work. The application of intensifying screens to increase the effective speed of photographic emulsions was an improvement made independently by numerous investigators.² But for quantitative measurements of intensity the photographic method is rather complicated, and it soon gave way to the ionization method, which is still used today.

The discharging effect of X-rays upon charged bodies, noted above, was soon traced to ionization of the air molecules. Villari³ showed

¹ COOLIDGE, *Phys. Rev.*, vol. 2, p. 409 (1913).

² *Comptes Rendus*, vol. 122, pp. 312, 702, 720 (1896); *Electrician*, vol. 36, p. 702 (1896).

³ VILLARI, *Comptes Rendus*, vol. 123, pp. 418, 446 (1896).

that the discharging action in a gas at a given pressure depends on the nature of the gas. The following were found increasingly active in the order given: H, CO, air, CO₂, ether vapor, CS₂. Benoist and Hurmuzescu¹ showed that, for a given gas, the discharging action increases rapidly with density.

At first, the rate of discharge of a charged electroscope was used in measuring the intensity of an X-ray beam, the motion of the leaves being observed under a low-power microscope. Later, an auxiliary device known as an "ionization chamber" was introduced. This is shown schematically in Fig. 142. *C* is a metal tube several centimeters in diameter, from about 20 to 100 cm. long, and closed at both ends except for an opening or "window" *W*, over which may be placed a thin sheet of cellophane or aluminum for admitting the

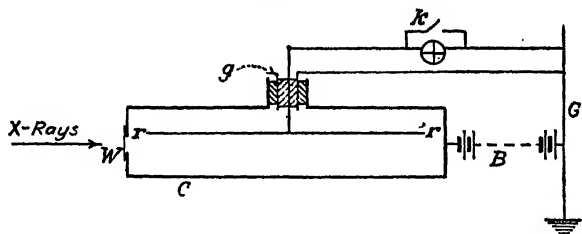


FIG. 142.—An ionization chamber used for measuring electrically the intensity of a beam of X-rays.

X-rays. A rod *rr* suitably supported by *good* insulating material, such as amber or quartz, is connected to an electrometer. An electric field is maintained between the rod *rr* and the cylinder *C* by a battery *B*, of about 100 volts, one end of which is connected to the ground wire *G*. An earthed guard ring *g* prevents leakage from the cylinder to the rod *rr*. The cylinder may be filled with a heavy gas to make the arrangement more sensitive; argon or methyl bromide is often used. When X-rays enter the window *W*, the gas within the cylinder is made conducting, and, on account of the electric field between the cylinder and the rod, the latter acquires a charge at a rate which can be measured by the electrometer. This rate is a measure of the intensity of the X-ray beam.

Nowadays the electrometer is often replaced by a vacuum-tube amplifier and a galvanometer. For very low X-ray intensities, a Geiger counter with appropriate vacuum-tube circuit [Sec. 209(a)] may supplant the ionization chamber.

183. Classical Pulse Theory of X-rays.—We have seen that X-rays are produced whenever electrons are suddenly brought to rest by colliding with a solid obstacle. This fact early suggested a simple

¹ *Comptes Rendus*, vol. 122, p. 926 (1896).

theory of the mode of their production. While being brought to rest, the electrons must experience, for a brief interval, a very large negative acceleration. According to classical electromagnetic theory, such an accelerated body must radiate energy. The sudden stopping of each electron as it collides with the target, therefore, must result in the emission of an electromagnetic disturbance or pulse. On this theory, X-rays consist of a very rapid succession of such pulses, coming at random intervals. It was shown by Stoney that, if such a stream of pulses is analyzed into wave trains, the components of shorter wave length are the more intense the greater the velocity of the electrons which are brought to rest by the target. Stoney showed also that matter should, in general, be more transparent to the shorter waves than to the longer waves. The hard, or penetrating, X-rays should, therefore, be produced by high voltages applied to the X-ray tube. Qualitatively, this picture of the mechanism of the production of X-rays seemed to be in agreement with the experimental facts.

(a) *The Production of X-rays.*—According to Eq. (39c) in Sec. 38, if an electron moving at a velocity v_0 , small relative to the velocity of light, is brought to rest with a uniform acceleration a , the total energy radiated is

$$W = \frac{2e^2v_0|a|}{3c^3},$$

e being the electronic charge. The time taken in stopping is $t = v_0/|a|$; during this time the electron travels a distance $s = |a|t^2/2 = v_0^2/(2|a|)$. In terms of s , we can also write

$$W = \frac{e^2v_0^3}{3c^3s}.$$

This formula predicts that the energy radiated will increase very rapidly with the velocity of the electron and will be inversely proportional to the distance within which the electron is stopped. Qualitatively, these same features might be expected to hold for stoppage of an electron by an atom, although in that case the acceleration would not, of course, be uniform. Experiment shows that the intensity of the radiation from a given X-ray target increases roughly as v_0^3 , i.e., as the square of the energy of the cathode ray, which is at least not inconsistent with our theoretical expectation. Furthermore, it is reasonable to assume that an electron moving with a given velocity will be stopped more suddenly when colliding with a heavy atom than with a lighter one. A target made of a heavy metal like platinum should, therefore, give out more X-ray energy, other things being equal, than one made of a lighter metal such as aluminum. This is in agreement with Roentgen's observations.

Qualitatively, therefore, the classical theory of X-ray production seemed to be well confirmed. Quantitatively, however, this theory was destined to meet with serious difficulties. It still retains an interest, although we now know that it has reference only to one part of the X-ray spectrum (known as the "continuous" spectrum) and, for accuracy, must be replaced even in dealing with that part by a more abstract theory, wave-mechanical in nature.

(b) *Refraction and Diffraction of X-rays*.—Tests of the classical pulse theory were by no means confined to predictions concerning the production of X-rays. Convincing demonstrations of such classical phenomena as reflection, refraction, and diffraction were eagerly sought. Roentgen's original experiments to this end had been negative.

Gouy,¹ using a narrow line source of X-rays, obtained a shadow of a fine platinum wire on a photographic plate. Then he placed a prism in the lower half of the beam in such a way that if the rays were *refracted* by the prism the shadow of the wire would be "broken." The shadow was sharp and continuous, showing no measurable refraction. Gouy estimated that the index of refraction could not be greater than 1.000005. It was pointed out by Maltezos² that the Helmholtz dispersion formula

$$n^2 = 1 + C\lambda^2 + \dots,$$

where n is the refractive index and C is a constant, predicts that $n \rightarrow 1$ as $\lambda \rightarrow 0$. Thus the absence of measurable refraction of X-rays would be accounted for if they had *extremely short wave lengths*.

Gouy³ also searched for evidences of diffraction, with negative results. A little later, however, Haga and Wind,⁴ using wedge-shaped slits only a few thousandths of a millimeter wide, observed evidences of a slight widening of the image on the photographic plate, from which they deduced that the wave length of the rays must be of the order of 10^{-8} cm. Walter and Pohl,⁵ in a similar diffraction experiment, confirmed this order of magnitude of the wave length.

184. The Scattering of X-rays.—In an attempt to discover the reflection of X-rays, Imbert and Bertin-Sans⁶ arranged apparatus

¹ Gouy, *Comptes Rendus*, vol. 122, p. 1197 (1896).

² *Comptes Rendus*, vol. 122, pp. 1115, 1474, 1533 (1896).

³ *Comptes Rendus*, vol. 123, p. 43 (1896).

⁴ *Ann. d. Physik*, vol. 68, p. 884 (1899).

⁵ *Ann. d. Physik*, vol. 29, p. 331 (1909). See also SOMMERFELD, *Ann. d. Physik*, vol. 38, p. 473 (1912).

⁶ *Comptes Rendus*, vol. 122, p. 524 (1896).

as shown diagrammatically in Fig. 143. Between the source S of the rays and the photographic plate P was placed a thick copper screen AA . A plane mirror M was so placed that a beam of rays, if reflected, would pass to the photographic plate P , on which would appear an image or shadow of an obstacle B . Such a shadow was obtained *irrespective of the angular position of the mirror M* . Indeed, a plate of paraffin was just as effective as the mirror. From these facts Imbert and Bertin-Sans concluded that, instead of being reflected, the rays were diffused or scattered by M , somewhat as light is scattered by fog particles.

This phenomenon of the scattering of X-rays has played a very important part in the theories of modern physics and has been the object of many researches both theoretical and experimental.

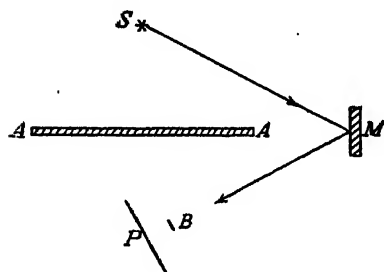


FIG. 143.—The arrangement of Imbert and Bertin-Sans showing pseudo-reflection (scattering) of X-rays from the mirror M .

(a) *Thomson's Theory*.—The first theory of X-ray scattering, an application of the classical pulse theory described in the last section, was proposed by Sir J. J. Thomson.¹ Let a beam of X-radiation be incident upon an electron of charge $-e$ and of mass m . The radiation, being supposed electromagnetic, contains an electric vector at right angles to the direction of propagation. As the

wave passes over the electron, the latter will experience a force $F = -Ee$ and will, therefore, have an acceleration in the direction opposite to F of magnitude

$$a = \frac{eE}{m}. \quad (261)$$

Consequently, the electron will radiate energy at the instantaneous rate given by Eq. (38) in Sec. 38, or

$$\Omega = \frac{2e^2a^2}{3c^3} = \frac{2e^4E^2}{3m^2c^3} \text{ ergs sec.}^{-1}$$

if e and E are in electrostatic units.

This energy is abstracted from the primary beam and is reradiated or *scattered* as a *secondary* beam. Its magnitude is most usefully expressed in terms of the intensity I of the primary beam. This intensity, defined as the amount of energy crossing unit area per

¹ THOMSON, J. J., "Conduction of Electricity through Gases," 2d ed., p. 321.

second, is

$$I = \frac{cE^2}{4\pi} \quad (261a)$$

[cf. Eq. (23a) in Sec. 34]. Hence,

$$\Omega = \frac{8\pi}{3} \frac{e^4}{m^2 c^4} I. \quad (262)$$

The ratio Ω/I , which we shall denote by σ_e , is called the *classical scattering coefficient for a free electron*. Of the incident radiation that falls on unit area of a surface drawn perpendicular to the beam, a fraction σ_e is scattered. Thus, we can say that the electron scatters as much radiation as falls on an area equal to σ_e . For this reason, σ_e is also called *the classical cross section for scattering by a free electron*. For its value, inserting $e = 4.80 \times 10^{-10}$, $m = 0.911 \times 10^{-27}$ gram, $c = 3 \times 10^{10}$ we find

$$\sigma_e = \frac{\Omega}{I} = \frac{8\pi}{3} \frac{e^4}{m^2 c^4} = 6.66 \times 10^{-25} \text{ cm.}^2 \quad (263)$$

(b) *The Number of Electrons in a Carbon Atom*.—The energy radiated by an electron must represent energy abstracted from the incident beam. If there are n electrons per unit volume in the scattering material, and if we assume that all electrons scatter independently (i.e., that no interference takes place between wavelets scattered from neighboring electrons), we may write for the fractional diminution $-\Delta I$ of the incident beam in going a distance Δx

$$-\frac{\Delta I}{I} = n\sigma_e \Delta x.$$

Since σ_e is independent of x , we may integrate both members of this equation, obtaining for the intensity I of the beam after traversing a thickness x of the material,

$$I = I_0 e^{-n\sigma_e x},$$

I_0 being the incident intensity. It is assumed here that the only loss of intensity is that due to scattering.

Actually, processes other than scattering are present and contribute to the diminution of the incident beam. Barkla and his collaborators,¹ however, were able to correct for the excess absorption, for absorbers of low atomic weight, and so were able to deduce, from measurements of I , I_0 , and x , a value of the product $n\sigma_e$. Using carbon of density ρ as the absorber, Barkla found

¹ BARKLA and SADLER, *Phil. Mag.*, vol. 17, p. 739 (1909); BARKLA, *Phil. Mag.*, vol. 21, p. 648 (1911).

$$n\sigma_e/\rho = 0.2,$$

from which and Eq. (263)

$$n = 3.00 \times 10^{23} \rho.$$

The number of atoms in a gram of carbon is N_0/A where N_0 is Avogadro's number or 6.02×10^{23} (Sec. 85), and A is the atomic weight (12); hence the number of atoms in 1 cm.³ is $N_0\rho/A$. Dividing this number into the value just found for n , we find for the number of electrons per atom of carbon

$$\frac{nA}{N_0\rho} = \frac{3.00 \times 10^{23} \times 12}{6.02 \times 10^{23}} = 6$$

(approximately). This was the first reasonably accurate determination of the number of electrons in an atom and was of great historical importance. However, it is only fair to say that, had Barkla performed his experiment under considerably different conditions of wave length and atomic number, the result of the computation would not have been so satisfactory.

(c) *Angular Distribution of Scattered X-rays.*—The distribution in direction of the radiation scattered by an electron is easily obtained from a formula deduced in Sec. 38. Consider the radiation that is scattered in a direction making an angle θ with the direction of the electric vector E in the primary beam. Let I_θ denote the intensity of the scattered radiation at a point P distant r cm. from the scattering electron in this direction. Then I_θ is the same as dW_1/dt as given by Eq. (36b) in Sec. 38, except that here $q = -e$; hence

$$I_\theta = \frac{1}{4\pi} \frac{e^2 a^2}{c^2 r^2} \sin^2 \theta,$$

or, after substituting again for the acceleration a in terms of the intensity I of the primary pulse by means of (261) and (261a),

$$I_\theta = \frac{I}{r^2} \frac{e^4}{m^2 c^4} \sin^2 \theta \text{ ergs cm.}^{-2} \text{ sec.}^{-1} \quad (264)$$

From this expression we see that if the primary X-ray beam is plane-polarized, with the electric vector in a fixed direction, then the radiation scattered by an electron in a direction making an angle θ with the direction of the primary electric vector E is proportional to $\sin^2 \theta$.

Let us consider now the more general case of a primary beam in which the direction of the electric vector varies at random. Such a

beam is unpolarized. Let us draw axes so that the xy plane contains the direction OP in which the scattered radiation is observed (Fig. 144), and let OP , which is perpendicular to the z -axis, make an angle ϕ with the x -axis or direction of propagation of the primary beam. Resolve the electric vector E in the primary beam into its y - and z -components, E_y and E_z . Then these components, acting upon the electron, will produce accelerations of magnitude $a_y = eE_y/m$ and $a_z = eE_z/m$, respectively. Each component of the acceleration, in turn, will produce a corresponding scattered electric vector at P . For these two scattered vectors we find, from Eq. (33) in Sec. 37, in which we insert $q = -e$ and, first $a = a_y$, $\theta = \pi/2 - \phi$, then $a = a_z$, $\theta = \pi/2$:

$$E'_1 = \frac{e^2 E_y}{c^2 m r} \cos \phi, \quad E'_2 = \frac{e^2 E_z}{c^2 m r}.$$

Since E'_1 and E'_2 are clearly perpendicular to each other at P , we have $E'^2 = E'^2_1 + E'^2_2$, where E' is the resultant electric vector in the scattered beam. If we insert the value of E'^2 thus found for E^2 in Eq. (23a) in Sec. 34, we obtain for the intensity of the scattered beam at P :

$$I'_\phi = \frac{c}{4\pi} E'^2 = \frac{1}{4\pi} \frac{e^4}{c^2 m^2 r^2} (E_y^2 \cos^2 \phi + E_z^2).$$

The average value of I'_ϕ can then be found by replacing E_y^2 and E_z^2 in this expression by their average values $\overline{E_y^2}$ and $\overline{E_z^2}$, the bar over a symbol denoting its time average.

The average intensity of the primary beam is, similarly,

$$I = \frac{c}{4\pi} \overline{E^2} = \frac{c}{4\pi} (\overline{E_y^2} + \overline{E_z^2}).$$

Because of the random variation of the direction of E , however, $\overline{E_y^2} = \overline{E_z^2}$. Hence,

$$\overline{E_y^2} = \overline{E_z^2} = \frac{2\pi I}{c}.$$

Thus, we can write for the average intensity of the scattered beam,

$$\overline{I}'_\phi = \frac{I}{r^2} \frac{e^4}{m^2 c^4} \frac{1 + \cos^2 \phi}{2}. \quad (265)$$

Assuming independent scattering by the individual electrons, we may multiply Eq. (265) by the number of electrons per unit volume, and thus we obtain an expression suitable for experimental test.

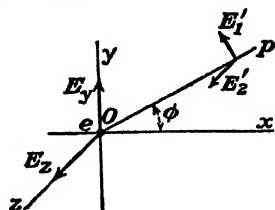


FIG. 144.

beam should be a maximum in the plane yOx_1 and should be zero in the direction $z_1z'_1$. If this analysis is correct, the intensity of the X-rays scattered by a substance at P_1 should vary from zero in a direction P_1z_1 to a maximum in the direction P_1P_2 .

In his experiments Barkla found that the intensity of the scattered radiation in the direction of P_1z_1 , although not zero, was considerably less than the intensity in the direction P_1P_2 . This indicates that the primary beam is at least *partially* polarized, though not completely so. A little further consideration indicates that this is just what we should expect. We assumed, at the beginning of the discussion, that the electrons of the cathode stream are brought to rest by accelerations in the direction Oz . If we attempt to picture the sequence of events by means of which a swiftly moving electron is brought to rest by collision with the atoms of the target at O , we should conclude that rarely will an electron be stopped by a single "head on" collision with an atom. In the general case, it will pursue a zigzag course and will collide with many atoms before being brought to rest. Although the preponderance of accelerations may be in the general direction Oz , accelerations in quite different directions are to be expected. In the primary beam, therefore, although the z -components of the electric vector should predominate, y -components are also to be expected. This means that the intensity of the X-rays scattered from P_1 in the direction of P_1P_2 should merely be greater than in the direction P_1z_1 —which is what Barkla observed.

With the *secondary* beam of X-rays proceeding in the direction P_1P_2 , however, the situation is different. This secondary beam is produced by the acceleration of electrons at P_1 due to the passage of the primary beam. On account of the transverse nature of electromagnetic radiation, this primary beam, regardless of its state of polarization, can accelerate electrons at P_1 only in directions lying in the plane $P_2P_1z_1$, i.e., in directions at right angles to P_1x_1 . Consequently, the electric vector of the *secondary* beam proceeding in the direction P_1P_2 must lie *entirely* in the plane $P_2P_1z_1$; this secondary beam must be *completely* plane-polarized. If, then, this secondary beam be allowed to pass over a second scatterer placed at P_2 , the intensity of the *tertiary* radiation sent out from P_2 should vary from zero in the direction P_2z_2 to a maximum in the direction P_2x_2 .

Barkla's experiments testing this conclusion indicated that the secondary rays were 70 percent polarized, instead of the predicted 100 percent. A similar experiment was performed much later by Compton and Hagenow¹ in which a more intense primary beam was

¹ COMPTON and HAGENOW, *J.O.S.A., Rev. Sci. Instruments*, vol. 8, p. 487 (1924).

employed, thereby allowing (1) better collimation, *i.e.*, greater definition of the scattering angle, and (2) the use of smaller scatterers so as to decrease multiple scattering, *i.e.*, successive scatterings at angles other than 90° . It was found that within the limit of error of the measurements the intensity of the tertiary radiation scattered in the direction P_{22} is zero. We may conclude, therefore, that, as classical theory predicts, the secondary beam itself is *completely polarized*.

The results of these early experiments on the scattering of X-rays confirmed without exception the classical theory in its simple form as described above. But we must bear in mind, that, as viewed today, this theory of scattering is imperfect, principally for the following reasons:

1. We have dealt here with free electrons only, whereas most of the electrons involved in scattering are bound in atoms.
2. We have neglected possible interference effects between wavelets scattered from different electrons.
3. Another type of scattering process, the Compton scattering, also occurs. In a later section we shall return to the subject of scattering.

185. Absorption and Fluorescence.—Although the classical electromagnetic theory was remarkably successful in explaining most of the early observed phenomena of X-rays, there gradually accumulated a mass of data which this theory could not explain. One such group of data developed on the subject of scattering, as already mentioned (Compton scattering), one in absorption (photoelectric absorption), and a third in regard to spectral distribution of energy of X-rays. The first work on the latter two groups of phenomena will be discussed in the last section dealing with early developments in X-rays.

(a) *Absorption.*—Roentgen's initial observation that the relative opacity of materials to X-rays depends not alone upon the density was early confirmed. Many investigators¹ independently concluded in 1896 that opacity of a material varied in a regular manner with its atomic weight. Not for some 15 years did the atomic number enter the picture.

In 1897, Buguet² made the significant observation that the opacity of a given material to X-rays varied with the thickness of the material previously traversed by the beam. This effect is explained by assuming that his X-ray beams were composed of several component beams

¹ See, for example, *Comptes Rendus*, vol. 122, pp. 146, 723 (1896).

² *Comptes Rendus*, vol. 125, p. 398 (1897).

having different absorbabilities. The general terms *quality* and *hardness* refer to the relative absorbability of a beam, a hard beam having low absorbability. Any beam whose quality is altered by transmission through an absorber is called a *heterogeneous* beam. In 1909, Barkla and Sadler,¹ studying these absorption questions, developed the first method of producing a *homogeneous* beam. This method is discussed in part (b) of this section.

The general formula for the absorption of radiation was found to apply to X-rays. That is, if an X-ray beam of intensity i is incident on a thin slab dx of absorbing material (Fig. 146), the change in intensity di , actually a diminution, is given by the relation

$$\frac{di}{i} = -\mu dx,$$

where μ is called the *linear coefficient of absorption*. For a homogeneous beam of X-rays, i.e., a beam for which μ is independent of x , the last equation can at once be integrated; we can write for the transmitted intensity I at any distance x

$$I = I_0 e^{-\mu x}, \quad (266)$$

I_0 being the intensity at $x = 0$. By measuring I , I_0 , and x , one may compute μ . In a later section, we shall discuss what is known concerning the dependence of μ upon atomic number and upon the quality of the rays.

In deriving Eq. (266), we are not concerned with what ultimately becomes of the radiation which is removed from the incident beam. From our discussion of scattered X-rays, however, we know that scattering constitutes one process by which an incident beam is weakened. Usually, however, the amount of radiant energy that is scattered from an X-ray beam is much smaller than the total amount of energy that is removed from the beam, as represented by the coefficient μ . This fact was not easy to explain on the classical theory.

(b) *Characteristic Secondary X-rays*.—A new type of phenomenon was revealed in 1896 by an important observation made by Winkelmann and Straubel.² A beam of X-rays was passed through a photographic plate P (Fig. 147), the emulsion being on the rear side. Behind a part of the plate was placed a piece of fluor spar F . On developing the plate, it was found that the film was much denser in

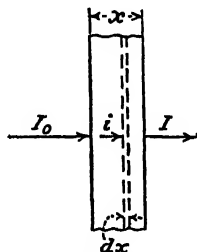


FIG. 146.

¹ *Phil. Mag.*, vol. 17, p. 739 (1909).

² *Jenaisch. Zeits. f. Naturwiss.*, vol. 30 (1896).

the neighborhood of F , as if F had reflected the rays. An observation of this kind had been made by Roentgen. But Winkelmann and Straubel showed that the phenomenon was not one of true reflection, for they repeated the experiment with a thin sheet of paper AA

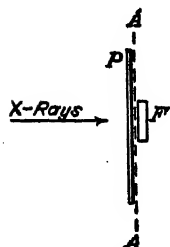


FIG. 147.

between F and P and found that the intensifying action of F was almost but not entirely destroyed, although the paper was very transparent to the *incident* beam. From this, they concluded that the *quality* of the rays had been altered by the spar in such a way as to make the beam of rays returned by F more absorbable in paper than was the original beam. In other words, the primary rays, incident on the spar, had been transformed into *characteristic* "spar" rays.

Barkla and his collaborators made a thorough and systematic study of this phenomenon.¹ Let a primary beam of hard X-rays from a target T (Fig. 148), after passing through holes in lead screens S_1S_1 , fall upon the secondary emitter E . Let the secondary beam, taken off at right angles to the primary beam, after passing screens S_2S_2 , enter the ionization chamber C , by means of which the intensity of the secondary beam can be measured either with or without slabs of absorbing material placed at A or at B .

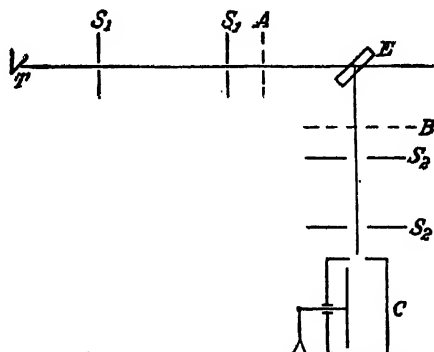


FIG. 148.—Schematic arrangement for studying secondary radiation.

When the secondary emitter E is of some light material, such as carbon, an aluminum absorbing screen placed at B absorbs nearly the same fraction of the secondary beam as it does of the primary beam when placed at A . This shows that the quality of the secondary beam as measured by its absorption in aluminum is nearly the same

¹ Cf. *Phil. Mag.*, vol. 16, p. 550; vol. 22, p. 396; vol. 23, p. 987; *Nature*, vol. 80, p. 37; *Cambridge Phil. Soc., Proc.*, vol. 15, p. 257 (1908–1912).

as that of the primary beam. We may conclude that the primary beam has merely been *scattered* by the secondary emitter.

If, however, a heavier material, such as silver, be substituted for the carbon at *E*, the absorption coefficient of the secondary beam in aluminum is found to be *greater* than that of the primary beam—which indicates that the *quality* of the secondary beam is no longer the same as that of the primary. Barkla found that the quality or hardness of the secondary beam, as measured by its coefficient of absorption in aluminum, is *characteristic of the material* used as secondary emitter. Almost without exception, *the hardness of the secondary beam, as measured by its absorbability in aluminum, increased with increasing atomic weight of the secondary emitter.*

When absorbing coefficients in different absorbing materials were compared, a striking feature appeared. If the absorption coefficients in, say, iron, for a series of secondary radiations obtained from a series of emitters of different atomic weights, are plotted against the absorption coefficients for the same radiations in aluminum, a curve is obtained of the type shown in Fig. 149. Point *a* corresponds to a soft radiation easily absorbed in aluminum. As the penetrating power of the radiation is increased, the coefficients of absorption in both aluminum and iron decrease along the line *ab*. At *b*, however, the absorption coefficient of the radiation in the iron absorber suddenly increases to point *c*. This increase occurs for radiation of which the absorbability in aluminum is about *the same* as that for the radiation from an iron secondary emitter. Thereafter, for more penetrating radiations, the absorption in iron decreases again toward point *d*.

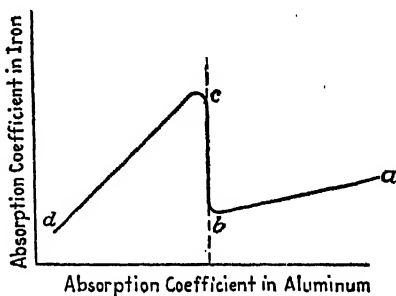


FIG. 149.

Similar results were obtained with other absorbing materials in place of iron. The position of the discontinuity was found to depend on the atomic weight of the absorber: the higher its atomic weight, the greater the penetration in aluminum of the radiation for which the discontinuity occurred.

These facts point unambiguously to the conclusion that a secondary emitter besides scattering the primary beam also emits a fluorescent radiation characteristic of the emitter whenever the primary beam is of greater hardness than this fluorescent radiation. Barkla called the latter the "*K fluorescent radiation*" of the emitter. For the heavier absorbers, a second discontinuity was observed beyond point *a*

of Fig. 149. This indicated that along with the *K* fluorescent radiation another and *softer* fluorescent radiation was emitted, which Barkla designated by *L*. Barkla recognized that these *K* and *L* radiations constitute *lines* (or groups of lines) in the fluorescent X-ray spectra of the several secondary emitters.

A further conclusion from these observations is that there exists, for each of many absorbers, a *K* absorption discontinuity which is not like the resonance absorption line to be expected from the classical electromagnetic theory. In some cases, a similar *L* discontinuity was found.

The graph of Fig. 149, with the abscissas corresponding, as we shall see, to an X-ray wave-length scale, foreshadows a new era in X-ray developments, an era distinguished by studies of spectral properties of X-rays and by experimentation with homogeneous beams.

X-RAY SPECTRA

186. The Crystal Diffraction Grating.—The researches of Barkla in isolating the *K* and *L* fluorescent radiations pointed unmistakably toward the existence of *definite, discrete* wave lengths in X-rays. An experiment was greatly needed by which a heterogeneous beam could be broken up into its spectral components and the discrete *K* and *L* radiations confirmed directly. The only method then available for such spectral analysis involved diffraction of X-rays from a slit, and the dispersion obtainable by this method was insufficient for the purpose at hand.

(a) *A Crystal as a Diffraction Grating.*—A new and practical method for resolving X-ray beams developed out of a brilliant suggestion by Laue. The order of magnitude of X-ray wave lengths, as revealed by the diffraction experiments described above, is the same as the order of magnitude of the spacing of the atoms in crystals. Laue suggested, therefore, that a crystal, with its regular three-dimensional array of atoms, might behave toward a beam of X-rays in somewhat the same way as does a ruled diffraction grating toward a beam of ordinary light.

Let it be assumed that plane electromagnetic waves traveling in a given direction fall upon a crystal. Then each atom will scatter some of the incident radiation. If the crystal is perfectly regular, the wavelets scattered by different atoms will combine, in general, in all sorts of phases and so will destroy each other by interference. Laue argued, however, that for certain wave lengths and in certain directions the wavelets should combine in phase and so produce a

strong diffracted beam.¹ It would be expected, therefore, that such diffracted beams might be observed upon passing a heterogeneous X-ray beam through a crystal. Such an experiment was performed by Friedrich and Knipping in 1913.

(b) *The Experiment of Friedrich, Knipping, and Laue.*²—By means of suitable screens, S_1S_2 (Fig. 150), a narrow pencil of X-rays from the target T was allowed to pass through a crystal C beyond which was a photographic plate PP . After an exposure of many hours, it was found on developing the plate that, in addition to the interior central image at O , where the direct beam struck the plate, there were present on the plate many fainter but regularly arranged spots, indicating that the incident X-ray beam had been diffracted by the crystal in certain special directions, just as Laue had predicted. Figure 151 shows such a photograph, taken by Dr. George L. Clark,³ of an iron crystal. In their original paper, Friedrich, Knipping, and Laue, from an analysis of a series of photographs

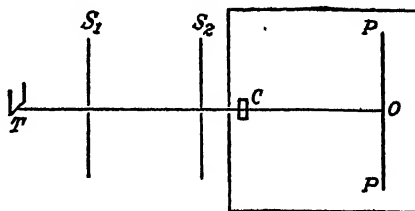


FIG. 150.—The arrangement by which Friedrich and Knipping discovered the action of a crystal on a beam of X-rays.

of a crystal of zinc blende oriented at various angles with respect to the incident pencil, concluded that there were present in the X-ray beam wave lengths varying between 1.27×10^{-9} cm. and 4.83×10^{-9} cm. This positive result proved the correctness of the *two* postulates underlying the experiment: (1) that X-rays are electromagnetic waves of definite wave lengths and (2) that the atoms of a crystal are arranged in regular three-dimensional order, as suggested by the external symmetry of crystals.

This experiment marked the beginning of a new era in the technique of X-ray measurement and in X-ray theory. Two new and very important fields of investigation were at once opened up: (1) in X-rays, the study of spectra and the use of homogeneous beams in experiments on scattering, absorption, etc.; (2) the study of the arrangements of atoms or molecules in crystals. In the following sections, we shall confine our discussion to some of the more important aspects of the former field. For presentations of the allied field of

¹ See, for example, COMPTON and ALLISON, "X-rays in Theory and Experiment," pp. 331-340, 1935.

² FRIEDRICH, KNIPPING, and LAUE, *Bayer. Akad. d. Wiss.*, 1912; *Le Radium*, vol. 10, p. 47 (1913).

³ This photograph is used by permission of Dr. Clark.

molecular and crystal structure, the student is referred to standard treatises on the subject.¹

(c) *Bragg's Analysis of the Three-Dimensional Grating.*—A very simple and convenient way of looking at the process of diffraction by a crystal grating was proposed by Bragg.² He pointed out that through any crystal a set of equidistant parallel planes can be drawn which, among them, pass through *all* the atoms (or similar groups of

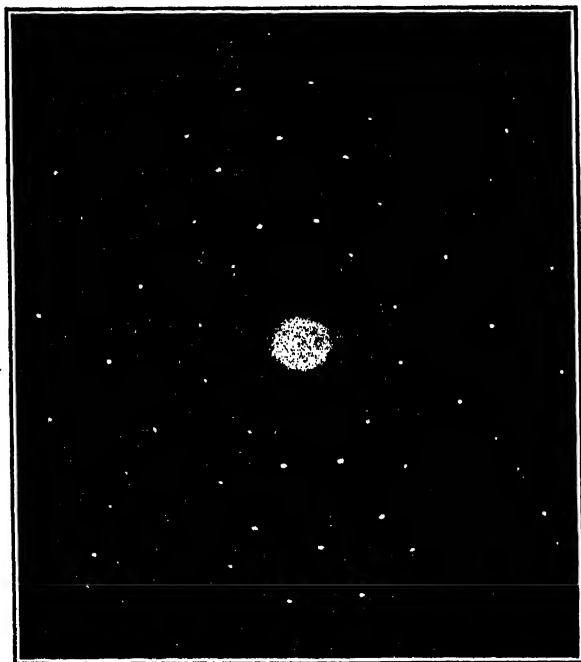


FIG. 151.—Laue photograph of an iron crystal. (Photograph by Dr. George L. Clark.)

atoms) which compose the crystal. Indeed, a great many such families of planes may be drawn, the planes of each family being separated from each other by a characteristic distance. Such planes are called *Bragg planes*, and their separations, *Bragg spacings*. Traces of five families of Bragg planes are shown in Fig. 152.

If plane monochromatic waves fall upon the atoms in a Bragg plane, a wavelet of scattered radiation spreads out from each atom in all directions. There is just one direction in which, irrespective of the atomic distribution in the plane, the scattered wavelets will meet

¹ *E.g.*, BRAGG, W. H., and W. L. BRAGG, "The Crystalline State," vol. I, 1933; CLARK, G. L., "Applied X-Rays," 3d ed., 1940.

² BRAGG, W. L., *Cambridge Phil. Soc., Proc.*, vol. 17, p. 43 (1912).

in the same phase and will constructively interfere with each other, *viz.*, the direction of *specular reflection* from the plane. This follows from the ordinary Huygens construction as used for the reflection from a mirror. The beam scattered in this direction may be thought of as reflected from the Bragg plane. But, now, we note that each Bragg plane is one of many regularly spaced parallel planes. The beams reflected from these various parallel planes will combine, in general,

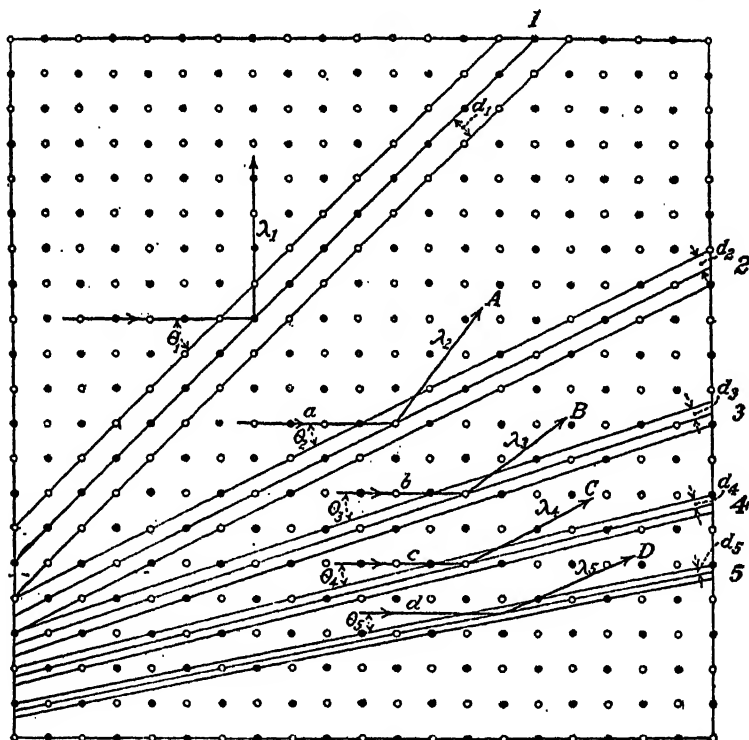


FIG. 152.—Schematic representation of the reflection of monochromatic beams of X-rays by a crystal of NaCl when a heterochromatic beam is incident upon it.

in different phases and so will destroy each other by interference. Only if certain conditions as to wave length and angle of incidence of the beam on the planes are satisfied will the waves from different planes combine in the same phase and reinforce each other. The necessary conditions are easily found.

In Fig. 153, let the horizontal lines represent the traces of two successive Bragg planes spaced d apart. Denote by θ the angle between the direction of propagation of the incident beam and the planes; this angle is called the *glancing angle* of the beam on these planes. Let a

ray meet the two planes at O_1 and O_2 , respectively, and let a line drawn from O_2 perpendicular to the planes cut the other plane at O . Draw O_1a and O_2b representing rays specularly reflected from the two planes, and draw aOb perpendicular to O_1a and O_2b to represent a wave front of the reflected beam. Then constructive interference will occur if the path O_1O_2b , taken by waves scattered at O_2 , exceeds the path O_1a for waves scattered at O_1 by an integral number of wave lengths. Draw Oc perpendicular to O_1O_2 . Then $O_1a = O_1c$; hence the differ-

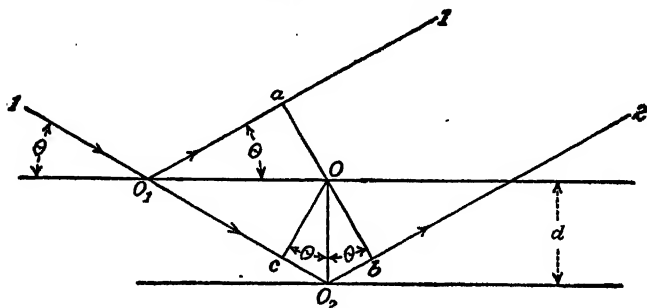


FIG. 153.

ence in path is $cO_2b = 2d \sin \theta$. The conditions that there should be a reflected beam are, therefore,

$$\theta = \theta' \quad (267)$$

and

$$n\lambda = 2d \sin \theta, \quad (268)$$

where n is an integer, called the "order of the reflection," and θ' is the angle between the Bragg planes and the direction of the diffracted rays. These two conditions are known as *Bragg's law* for X-ray reflection. The first condition is often omitted in the statement of Bragg's law with the understanding that it is implied by the term "reflection."

Suppose, now, that a parallel wave train, containing a *continuous spectrum* of wave lengths, is incident upon a crystal, as represented by the parallel arrows a, b, c, d in Fig. 152. In the figure, traces of five families of Bragg planes are shown, numbered 1, 2, 3, 4, 5, with their characteristic spacings d_1, d_2, \dots . Many other families of planes might be imagined, some perpendicular and some not perpendicular to the plane of the paper. Suppose that in the incident beam there is a wave length λ_2 such that

$$n\lambda_2 = 2d_2 \sin \theta_2,$$

where n is an integer, d_2 is the distance between the set of planes numbered "2," and θ_2 is the glancing angle between the direction of

the incident radiation and these planes. Then there will be reflected from this group of planes a beam A , of wave length λ , which will proceed in the direction of the arrow A . Similarly, we may have reflected beams B, C, D, \dots in different directions in the plane of the paper, and also many other beams reflected from other families of planes in directions not in the plane of the paper. Each "Laue spot" in the experiment of Friedrich and Knipping may be interpreted as produced by such a reflected beam. In general, the most intense spots correspond to reflections from Bragg planes containing the greatest number of atoms on each plane.

The crystal represented in Fig. 152 is one of a very simple type, but the conception of Bragg planes is applicable to all types of crystals. It should be pointed out, however, that it is not necessary to draw the Bragg planes actually *through* the atoms; instead of a given family of these planes, any other set of planes parallel to them and spaced the same distance apart could be employed and would lead to the same conditions for strong reflection.

It should be pointed out also that the Bragg equation (268) does not give a complete solution to the interference problems of X-rays scattered from a crystal. The equation predicts only the position of the center of the expected diffraction pattern for a given wave length and family of Bragg planes; nothing is said about the intensity distribution in this diffraction pattern.

187. The X-ray Spectrometer.—Immediately following the announcement by Friedrich, Knipping, and Laue of their successful experiment, many investigators took up a study of the new phenomenon. Among these were W. H. and W. L. Bragg,¹ to whom we are chiefly indebted for the early development of the X-ray spectrometer.

A spectrometer of the Bragg type is shown diagrammatically in Fig. 154(a). X-rays from the target T of an X-ray tube pass through two narrow slits S_1 and S_2 , a few hundredths or tenths of a millimeter wide, the edges of which are made of some material, such as lead or gold, which is very opaque to X-rays. This ribbon-shaped incident beam of X-rays I falls at a glancing angle θ on the cleavage face of a crystal K —rock salt, calcite, mica, gypsum, quartz, etc.—which is mounted on a table D , the angular position of which can be accurately read by verniers or micrometer microscopes. The reflected beam of X-rays, which makes an angle 2θ with the incident beam, enters, through the "window" w , an ionization chamber C by means of which the intensity of the reflected beam may be measured. By suitably

¹ BRAGG, W. L., *Nature*, vol. 90, p. 410 (1912); BRAGG, W. H., and W. L. BRAGG, *Roy. Soc., Proc.*, vol. 88, p. 428 (1913); BRAGG, W. H., *Nature*, vol. 91, p. 477 (1913).

turning the table D about the axis A , the incident beam may be made to strike the face of the crystal at any glancing angle θ . The ionization chamber C is mounted on an arm (not shown) by means of which the chamber can be rotated about the axis A so as to admit the reflected beam through the window w . For protection against stray scattered radiation, a third slit S_3 is attached to the chamber.

For photographic registration, the ionization chamber may be replaced by a photographic plate PP [Fig. 154(b)]. With the crystal

set at a glancing angle θ , the reflected beam will strike the plate at L (or at L' , if the crystal be "reversed"). From the position O at which the direct beam strikes the plate, the distances OL and OA and hence the angles 2θ and θ may be determined. The wave length λ is then obtained from the Bragg formula, $n\lambda = 2d \sin \theta$.

The distance d between the reflecting planes of a crystal such as NaCl is determined as follows. From his investigations, Bragg showed¹ that, in the rock-salt crystal, the Na and the Cl atoms or ions occupy alternate positions at the corners of elementary cubes in the cubic lattice characteristic of the crystal, the arrangement

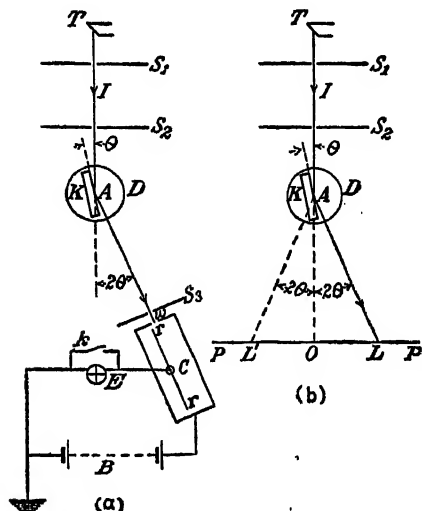


FIG. 154.—The X-ray spectrometer, using (a) the ionization method, (b) the photographic method.

being similar to that shown in Fig. 152, which represents one plane of atoms. Taking the atomic weight of chlorine as 35.46 and of sodium as 23.00, we find the molecular weight of NaCl to be 58.46. Therefore, 58.46 grams of the NaCl contain $2N_0$ atoms, *i.e.*, N_0 atoms of Na and N_0 atoms of Cl , where N_0 is Avogadro's number. Thus, if we use the older value, $N_0 = 6.064 \times 10^{23}$, we find for the number of atoms n in 1 cm.-cube of rock salt

$$n = 2 \times 6.064 \times 10^{23} \times \frac{\rho}{58.46},$$

where $\rho = 2.163$, the density of NaCl . If d is the distance between the center of one atom and the next along the edge of the cube, $1/d$ is the number of atoms in a row of atoms 1 cm. long, and the number

¹ See BRAGG, W. H., and W. L. BRAGG, "The Crystalline State," vol. I, 1933.

of atoms in the centimeter-cube is

$$n = \frac{1}{d^3}.$$

Solving these two equations simultaneously for d , which is the desired distance between the (cleavage) planes in NaCl, we find

$$d = 2.814 \times 10^{-8} \text{ cm. (Old value.)}$$

The value of d is seen to be dependent on M , the molecular weight of rock salt; on ρ , the density of rock salt; and on Avogadro's number, N_0 . The values of N_0 and ρ are not known to much better than 0.1 percent. There is, hence, a corresponding uncertainty in the value of d . At present, measurements of X-ray wave lengths can be made with a precision manifold greater than the best measurements of M , N_0 , or ρ .

It was, therefore, supposed to be expedient to adopt a new unit of length, called the "X unit" or "X.U.," which was *very nearly* 1×10^{-11} cm. but which was accurately defined by taking the grating space of NaCl at 18°C. as *exactly* 2,814.00 X.U. The grating space of any other crystal could then be determined from this arbitrarily chosen standard as follows. Let θ_{NaCl} be the angle at which a line of given wave length is reflected from a rock-salt crystal, and θ_c the angle at which the same line is reflected from some other crystal of grating space d_c . Then, from Bragg's law [Eq. (268)], we have

$$n\lambda = 2d_{\text{NaCl}}\sin\theta_{\text{NaCl}} = 2d_c\sin\theta_c,$$

from which d_c may be determined. In later work, calcite was adopted as the best crystal for practical use, but the standard unit for wave lengths was not changed.

Recently, however, it has been found that there was an error in the old value of e (Sec. 45) and in the value of N_0 deduced from it. If the present value, $N_0 = 6.023 \times 10^{23}$, is used in the preceding calculation, we find

$$d = 2.820 \times 10^{-8} \text{ cm. (New value.)}$$

Furthermore, it has been found possible to measure X-ray wave lengths by means of ruled gratings, the grating space of which can be measured directly, so that no assumptions as to structure of a crystal or as to the value of N_0 are involved. It seems certain that in the future the absolute values of X-ray wave lengths will be based upon measurements with ruled gratings, crystals being used only to compare one wave length with another.

Under these circumstances the old definition of the X unit loses all its advantage, and it is preferable to redefine the X unit as being *exactly* 10^{-11} cm. The latter definition will be assumed hereafter.

There are two important corrections which must be made when, by use of a crystal grating, X-ray wave lengths are measured with a precision of the order of a few parts in 100,000. First, because of the thermal expansion of crystals, the grating space d varies with temperature, and correction must be made if the temperature differs from 18°C. , which is the temperature to which grating spaces are usually referred. Second, it has been found that a beam of X-rays is slightly deviated by refraction as it enters or leaves a crystal (see Sec. 199 below). Accordingly the *observed* value of the glancing angle θ of the Bragg formula is not quite the same as the angle, *within* the crystal, at which the beam of rays strikes the Bragg planes.

In Table I are listed some of the crystals commonly used in X-ray spectroscopy together with their respective grating spaces. In the second column, the spaces are listed as given by Siegbahn,¹ on the basis of $d = 2,814$ X.U. for rock salt; in the third column, headed "corrected," the spaces are given as found by multiplying Siegbahn's values by $3.03560/3.02945$ so as to bring them into harmony with the grating spacing of calcite ($d = 3,035.60$ X.U.) as calculated by Bearden from observations of the diffraction by a calcite crystal of X-rays whose wave length had been measured with a ruled grating.²

TABLE I.—GRATING SPACES OF SOME CRYSTALS USED IN X-RAY SPECTROSCOPY

Crystal	Grating space d at 18°C. , X.U.		Change in d per degree centigrade, X.U.
	Siegbahn	Corrected	
Rock salt (NaCl).....	2,814.00	2,819.71	0.11
Calcite (CaCO_3).....	3,029.45	3,035.60	.03
Quartz (SiO_2).....	4,246.02	4,254.65	.04
Gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$).....	7,584.70	7,600.1	.29
Mica.....	9,942.72	9,962.9	.15

It is quite beyond the scope of this book to discuss the various (ionization) spectrometers or (photographic) spectrographs which have been developed for the measurement of X-rays. Some, *e.g.*, the Bragg spectrometer, use a single plane crystal and "external"

¹ SIEGBAHN, "Spektroskopie der Röntgenstrahlen," 2d ed., 1931.

² BEARDEN, *Phys. Rev.*, vol. 48, p. 385 (1935).

reflection, some use transmission and "internal" reflection, some use two plane crystals in succession, and some use a bent crystal in order to focus the reflected rays. For details of the various designs and of their operation, the reader is referred to other books.¹

It is obvious from Bragg's law,

$$n\lambda = 2d \sin \theta,$$

that the maximum wave length λ_m measurable by use of a given crystal is $2d$. Actually, λ_m is somewhat less than this, since the glancing angle

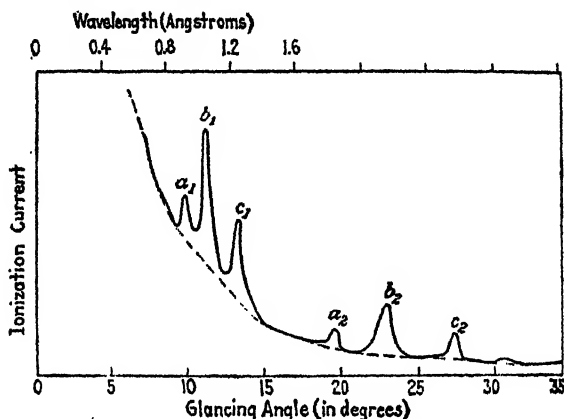


FIG. 155.—Bragg's curve for the energy distribution in an X-ray spectrum, showing the characteristic lines a , b , c .

θ cannot usefully exceed about 70° . By use of a quartz crystal, for example, wave lengths up to about 7,500 X.U. can be measured. For measurement of extremely long wave lengths, certain organic crystals of large grating space were originally used (*e.g.*, sugar, $d = 10.57$ Å; lead melissate, $d = 87.5$ Å). In recent years, however, it has been found more expedient to use a ruled grating for this purpose. There is no such upper limit to the wave lengths which may be measured by means of a ruled grating. In general, crystals are now used from very short wave lengths up to about 50 Å, and gratings from very long wave lengths to somewhat below 50 Å.

188. Bragg's Discovery of Monochromatic Characteristic Radiations.—With a beam of X-rays from a platinum target incident on the cleavage face of a rock-salt crystal, as shown in Fig. 154(a), W. H. Bragg² rotated the crystal in steps of $\Delta\theta$ and the ionization chamber

¹ SIEGBAHN, *op. cit.* See particularly the two-crystal spectrometer, pp. 128-134, an instrument of very high resolving power. Cf. also, Compton and Allison, *op. cit.*, pp. 750-756 for focusing spectrographs.

² *Nature*, Jan. 23, 1913.

in steps of $2\Delta\theta$. He plotted the curve of ionization current against glancing angle θ and found that the current or X-ray intensity did not vary uniformly with angle but rose at certain angles to a sharp maximum. A curve similar to that shown in Fig. 155 was obtained. A group of three maxima, a_1 , b_1 , and c_1 , was observed at the respective angles θ of 9.9, 11.6, and 13.6°. A second group of three maxima, a_2 , b_2 , and c_2 , was observed at approximately double these angles. This second group is similar, as to relative intensities of the maxima, to the first group. Bragg interpreted the maxima a_1 , b_1 , and c_1 as three monochromatic lines; and the second group of maxima a_2 , b_2 , and c_2 as second-order reflections of the lines a_1 , b_1 , and c_1 . He computed their wave lengths by the formula $n\lambda = 2d \sin \theta$, taking $n = 1$ for the "lines" a_1 , b_1 , and c_1 and $n = 2$ for the second-order lines a_2 , b_2 , and c_2 . A third-order peak b_3 was observed at 36.6°. Assuming $d = 2.814$ angstroms for rock salt, he found the following wave lengths:

TABLE II.—FIRST MEASUREMENTS OF X-RAY WAVE LENGTHS BY BRAGG
(PLATINUM TARGET)

Line	θ , degrees	$\sin \theta$	n	λ , angstroms
a_1	9.9	0.172	1	0.97
b_1	11.6	.200	1	1.13
b_2	23.6	.400	2	1.13
b_3	36.6	.597	3	1.12
c_1	13.6	.235	1	1.32

Curves similar to Fig. 155 were obtained with other crystals, the only difference being that the maxima occurred at different glancing angles, indicating that each crystal had a characteristic grating space d . Bragg convinced himself, however, that these respective maxima for different crystals always represented the same monochromatic radiation, since, for example, the absorption in aluminum of peak b_1 was always the same, whatever the crystal used. In short, the peaks of the curve in Fig. 155 represent *spectral lines* the wave lengths of which are *characteristic of the target emitting the rays*. These monochromatic lines are superimposed on a *continuous* spectrum represented by the partially dotted line in the figure. Curves of the type shown in Fig. 155, therefore, represent (subject to certain corrections to be mentioned later) the distribution of *energy* in the X-ray spectrum, continuous and characteristic combined, of an element.

189. Moseley's Law.—In two classic papers,¹ Moseley presented a systematic study of the characteristic radiations emitted by various

¹ MOSELEY, *Phil. Mag.*, vol. 26, p. 1024 (1913); vol. 27, p. 703 (1914).

targets, using a photographic method similar in principle to that shown in Fig. 154(b). He found a larger number of characteristic lines than Bragg and, also, that these lines could, in general, be classified into two groups: (1) a group of shorter wave lengths, which, by means of the value of absorption coefficients in aluminum, he identified with Barkla's *K* characteristic secondary radiations and (2) a group of lines of longer wave length, similarly identified with Barkla's *L* radiation.

Unlike the optical spectra, however, the X-ray characteristic spectra of the elements were found to be similar from element to

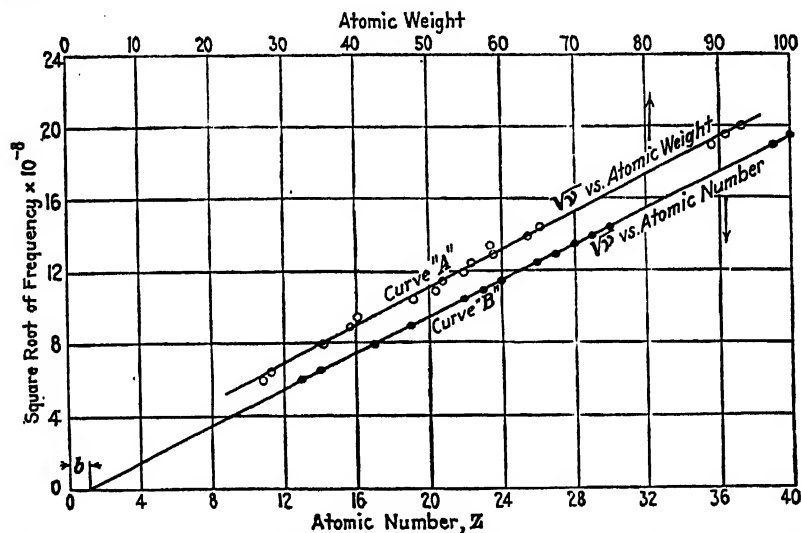


FIG. 156.—Moseley's curve showing the relation between the frequencies of X-ray lines and atomic number.

element, homologous lines occurring, in general, at shorter wave lengths the greater the atomic weight of the element in which the lines originate. In searching for a relation between the frequency of a given line, say the *K* α line, and some property of the atom in which the line originated, Moseley first observed that the frequency did not vary uniformly with the atomic weight, as is shown by curve A (Fig. 156) in which the square root of the frequency is plotted against the atomic weight. In Bohr's theory of the origin of spectra, however, which had recently been proposed, the charge on the nucleus played a fundamental role. According to Bohr's theory, the frequency ν of a spectral line is given by

$$\nu = Z^2 \frac{2\pi^2 mc^4}{h^3} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right)$$

[Sec. 104, Eq. (140a)], from which

$$\sqrt{\nu} \propto Z,$$

Z being the atomic number. Rutherford had shown, from his experiments on the scattering of α particles, that the value of the nuclear charge, for a given atom, is very approximately one-half the atomic weight; and Barkla had shown, from experiments on the scattering of X-rays, that the number of electrons surrounding the nucleus is also approximately one-half the atomic weight. Guided by these considerations, Moseley assigned atomic numbers Z to the elements he had investigated and then plotted a curve between $\sqrt{\nu}$ and Z . Such a plot of Moseley's data for the $K\alpha$ line is shown in Fig. 156, curve B . The graph is seen to be a straight line, with a small intercept $b = 1$ on the Z -axis. It is obvious from a comparison of the two curves A and B of Fig. 156 that, as far as the determination of the frequency of characteristic lines is concerned, atomic number is a much more fundamental quantity than atomic weight.

Empirically, the relation between the frequency ν of the $K\alpha$ line and Z , as determined by Moseley from Fig. 156, curve B , is

$$\nu = 0.248 \times 10^{16} (Z - 1)^2. \quad (269)$$

In Bohr's equation for the frequency ν of a spectral line, as just written, if we set $n_2 = 1$ and $n_1 = 2$, we obtain for ν

$$\nu = 0.328 \times 10^{16} \times Z^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right) \quad (270a)$$

$$= 0.246 \times 10^{16} \times Z^2 \quad (270b)$$

by inserting the numerical values of m , e , and h . Except for the slight correction to Z , Eqs. (269) and (270b) are seen to be almost identical.

This agreement suggests the hypothesis described in the next section as to origin of the $K\alpha$ line.

190. The Origin of X-ray Lines. (a) *Emission Spectra.*—In Sec. 128, we have seen that in an atom there are 2, and only 2, electrons corresponding to the electronic quantum number, $n = 1$. These electrons, which constitute the K shell and may be called the K electrons, have 1-electron wave functions (in the zero-order stage of perturbation theory) the values of which are very small except close to the nucleus. The electrons in question are, therefore, very unlikely to be found at any considerable distance from the nucleus and may be regarded as the innermost electrons in the atom. In any neutral

atom of atomic number $Z > 2$, there exist also 1 or more, up to a maximum of 8, electrons represented by wave functions with $n = 2$, called " L electrons." Their wave functions extend effectively to much greater distances from the nucleus than those of the K electrons, so that, although any one of the L electrons might be found closer to the nucleus than the K electrons, they are most likely to be found at much greater distances. Then, if Z is large enough, there may occur, also, M electrons, with $n = 3$, characterized by wave functions extending still farther out, and so on, until all of the circumnuclear electrons are accounted for.

Now suppose an atom in the target of an X-ray tube is bombarded by an energetic cathode ray and that 1 of the 2 K electrons is "knocked out" of the atom. The atom is thereby converted into an ion, and the ion is left in a quantum state characterized by the absence of 1 K electron; this state may be called a *K quantum state of the ion*. Subsequently an L electron may "drop" into the K vacancy, i.e., change from a 1-electron wave function with $n = 2$ to one with $n = 1$; the atom thereby undergoes a transition from a K state to an L state, its excess energy being emitted as a quantum of radiant energy. The emission of the $K\alpha$ line is ascribed to atomic processes of this sort. The quantum $h\nu$ of $K\alpha$ radiant energy is thus equal to the difference in energy between the K and the L quantum states or levels of the ionized atom.

In a similar way, the $K\beta$ line may be assumed to originate when an M electron, with quantum number $n = 3$, drops into the K vacancy; i.e., when there occurs an atomic (ionic) transition from a K state to an M state. The picture may be extended to account for, or to predict, any one of a large number of characteristic emission lines. A line may be expected corresponding to any atomic (ionic) transition from an initial state characterized by the absence of an electron of quantum number n to a final state corresponding to the absence of any electron of quantum number greater than n . Of course, in an actual target in an X-ray tube many atoms are simultaneously involved and many lines are emitted simultaneously.

This picture suggests, also, an explanation, at least qualitative, of the appearance of the factor $(Z - 1)$ instead of Z in Moseley's equation, (269). When, as a result of bombardment by the cathode stream, 1 of the K electrons is removed from an atom, there is 1 K electron left near the nucleus. This electron "screens" the nucleus and makes its *effective* nuclear charge about 1 unit less; hence, the factor $(Z - 1)$. Further work has shown, however, that Moseley's law holds only as a first approximation.

After the work of Bragg and Moseley, the techniques of X-ray spectrometry developed rapidly. Many new characteristic lines were discovered, and at the present time it is believed that practically the entire spectrum of most elements has been mapped out. Following

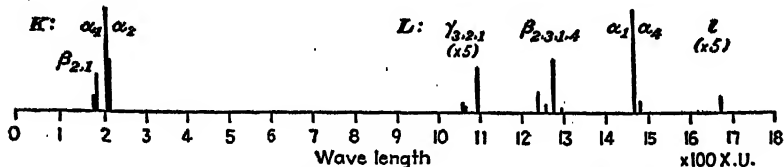


FIG. 157.—Plot of the principal K and L lines of tungsten against wave length. The relative intensity of the lines, when observed under certain conditions, is indicated roughly by the heights of the lines on the plot, the heights for the $L\gamma$ and l lines being increased in the ratio 1:5.

and extending the notation of Barkla, the lines are classified into series known as the K , L , M , N , etc. series, respectively; the letter designating a series refers to the initial atomic (ionic) state in the transitions giving rise to the lines in question. The principal lines in the K series and L series from tungsten are plotted in Fig. 157, as explained

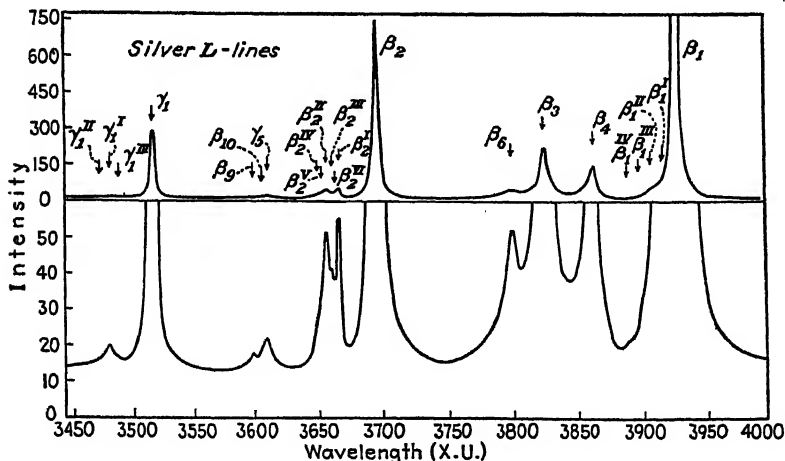


FIG. 158.—Distribution of intensity on an arbitrary scale, as measured with an ionization chamber from a target of silver, in the $L\beta$ region. The three groups of lines, β'_1 to $\beta_{(IV)}$, β'_2 to $\beta_{(VI)}$, and γ'_1 to $\gamma_{(VII)}$, are satellite lines [Sec. 201(a)]. [From L. G. Parratt, *Phys. Rev.*, vol. 54, p. 99 (1938)].

under the figure. Each series of lines contains several strong lines and numerous faint ones. To illustrate further the appearance of X-ray spectra, there is shown in Fig. 158 a section of the L series of silver, which was recorded with a two-crystal ionization spectrometer of high resolving power and dispersion.

Many faint X-ray lines, called "satellite lines," originate in atomic transitions between states of double ionization and will be discussed in more detail in a later section [201(a)]. Except in that section, the

TABLE III(a).—WAVE LENGTHS IN X UNITS OF THE MORE PROMINENT *K* LINES OF SOME REPRESENTATIVE ELEMENTS*

Element	<i>Z</i>	γ^\dagger	β^\dagger	α_1	α_2	λ_K
S	16	5,021.1	5,361.3	5,363.7	5,008.8
Ca	20	3,083.4	3,351.69	3,354.95	3,064.3
Fe	26	1,753.01	1,932.08	1,936.01	1,739.4
Zn	30	1,281.07	1,292.55	1,432.17	1,436.03	1,280.5
Br	35	918.53	930.87	1,037.59	1,041.66	918.09
Mo	42	619.70	630.98	707.83	712.11	618.48
Ag	47	486.30	496.01	558.28	562.67	484.48
I	53	374.71	383.15	432.49	437.03	373.44
W	74	178.98	184.22	208.62	213.45	178.22
Pb	82	141.25	146.06	165.16	170.04	140.49
U	92	108.42	111.87	126.40	130.95	106.58

* See Note 1 below.

† See Note 2 below.

TABLE III(b).— ν/R VALUES OF THE MORE PROMINENT *K* LINES OF SOME REPRESENTATIVE ELEMENTS*

ν = frequency; R = Rydberg constant

Element	<i>Z</i>	γ^\dagger	β^\dagger	α_1	α_2	$(\nu/R)_K$
S	16	181.49	169.97	169.90	181.93
Ca	20	295.54	271.88	271.62	297.38
Fe	26	519.83	471.65	470.69	523.90
Zn	30	711.34	705.02	636.29	634.58	711.67
Br	35	992.10	978.95	878.25	874.82	992.57
Mo	42	1,470.52	1,444.23	1,287.42	1,279.69	1,473.4
Ag	47	1,874.96	1,837.20	1,632.29	1,619.55	1,880.9
I	53	2,431.9	2,378.4	2,107.0	2,085.2	2,440.2
W	74	5,091.4	4,946.6	4,368.2	4,269.3	5,113
Pb	82	6,452	6,239	5,518	5,359	6,486
U	92	8,405	8,146	7,209	6,959	8,550

* See Note 1 below.

† See Note 2 below.

discussion will be confined to "first-order lines," arising in the manner just described from transitions between states of *single* ionization.

In Table III(a) are listed the wave lengths of the four principal lines of the *K* series, the α_1 , α_2 , β , and γ lines, for a number of elements.

TABLE III(c).— ν/R VALUES OF THE L ABSORPTION LIMITS OF CERTAIN ELEMENTS*

Element	Z	L_I	L_{II}	L_{III}
Mo	42	212.43	193.39	185.81
Ag	47	280.85	259.91	246.76
I	53	382.26	357.71	335.78
W	74	890.37	850.28	715.30
Pb	82	1,166.47	1,119.05	960.03
U	92	1,604.3	1,541.0	1,264.2

* See Note 1 below.

TABLE III(d).— ν/R VALUES OF THE M ABSORPTION LIMITS OF CERTAIN ELEMENTS*

Element	Z	M_I	M_{II}	M_{III}	M_{IV}	M_V
W	74	208.77	189.95	167.9	140.5	136.0
Pb	82	283.7	262.1	225.9	192.0	184.3
U	92	408.9	382.1	316.7	273.9	261.0

* See Note 1.

Note 1.—The wave lengths and ν/R values in these tables are taken from Siegbahn.¹ To bring them into harmony with modern wave-length measurements (Secs. 187, 200), they must all be increased by 0.20 per cent.

Note 2.—The $K\beta$ line is really a close doublet, experimentally resolved for Rb 37 and elements of higher atomic number. The two components are called (Siegbahn) β_1 and β_2 , β_1 being of shorter wave length. For Ag 47, for example, the wave lengths are $\lambda_{\beta_1} = 496.01$ X.U.; $\lambda_{\beta_2} = 496.65$ X.U. In the table, the wave length of β_1 is given for elements for which the line is resolved. The γ line is also a close doublet.²

It is more convenient from the theoretical standpoint, however, to work with quantities proportional to the frequencies of the lines instead of with their wave lengths; for this purpose values of ν/R are commonly employed, ν being the frequency and R denoting the Rydberg constant (in the same units as ν). Values of ν/R for the same K lines are shown in Table III(b).

It is to be noted that the phrase "series of lines" as used in X-ray spectroscopy refers to a group of lines arising from a common *initial* atomic state, whereas in the part of spectroscopy that deals with the outermost electrons of the atom the term series is applied to a group of lines having a common *final* atomic state.

(b) *Absorption Spectra*.—The theoretical picture that we have described furnishes, furthermore, a simple explanation of the dis-

¹ *Loc. cit.*² See HUDSON and VOGT, *Nat. Acad. Sci., Proc.*, vol. 19, p. 447 (April, 1933).

continuities that Barkla had observed in the absorption coefficient of various materials for X-rays. Absorption, as we saw in studying atomic spectra, is the reverse of emission. If a photon of $K\alpha$ radiation is emitted when an electron drops from an L shell into a K shell, then it should be possible for a photon of this frequency to be absorbed by an atom while one of the K electrons is raised into an L shell—provided there is a vacancy in the L shell into which it can go. As we have seen, however, in any atom with $Z > 10$, the L shell normally contains as many L electrons as can get into it. Hence for such atoms the $K\alpha$ lines cannot actually be observed as absorption lines. The same is true, for analogous reasons, of all of the X-ray emission lines that are commonly observed. A photon *can* be absorbed, however, if it has enough energy to remove an inner electron from the atom entirely. Let W_K denote the energy required to remove a K electron and to leave it at rest outside the atom. Then a photon of frequency ν can eject a K electron provided $\nu \geq \nu_K$ where

$$h\nu_K = W_K.$$

The explanation of Barkla's absorption curves is now clear. As ν is progressively increased, at the frequency $\nu = \nu_K$ the absorption suddenly increases, *i.e.*, absorption in the K shell begins; and thereafter it continues for all larger values of ν . The curve representing the absorption coefficient plotted against frequency or wave length will, therefore, show a sudden rise toward the side of shorter wave length at $\nu = \nu_K$. This is what Barkla observed, as is illustrated in Fig. 149, where the abscissa can also be regarded as representing wave lengths plotted on a certain scale, with shorter wave lengths toward the left. The critical wave length λ_K , or frequency ν_K , at which absorption in the K shell begins is called the *K absorption limit*. Values of λ_K and of ν_K/R are listed in the last columns of Tables III(a), (b). Since, in the theoretical picture, the energy W_K required to remove a K electron is somewhat greater than the energy change when the atom undergoes a transition between its K state and another of its discrete quantum states, it is to be expected that $h\nu_K$ will be greater than the corresponding quantity for any of the K lines. Hence ν_K should exceed ν for any of these lines, but all of these frequencies should be of the same order of magnitude. These conclusions are in agreement with the observations of Barkla.

The intimate connection between the emission lines of the K series and the K absorption limit is further shown by the critical voltage which must be applied to the X-ray tube in order to generate these lines. The critical voltage V_K is found to be determined by the

relation

$$eV_K = h\nu_K,$$

where ν_K is the frequency of the K absorption limit for the material of which the target is composed, h is Planck's constant, and e is the (numerical) electronic charge. Furthermore, *all of the lines of the K series are excited at the same critical voltage V_K* , and, as the voltage is increased above V_K , these lines all increase in intensity at exactly the same rate. In the case of tungsten, for example, for which $\lambda_K = 178.22$ X.U., we find from the last equation, or from Eq. (147b) in Sec. 108, that $V_K = 69,300$ volts. When voltages higher than this are applied to a tungsten-target X-ray tube, all of the K -series lines of tungsten appear. The explanation of these facts, in terms of the picture described above, is obvious.

Similar facts and interpretations hold for each of the other series of characteristic X-ray lines, the L , M , N , . . . series. A significant difference, however, is that here more than one absorption limit exists in each case. Thus, *three L absorption limits are found*, denoted, in order of decreasing frequency, by L_I , L_{II} , and L_{III} . There are *five M limits, seven N limits*, etc. The theoretical reason for the existence of these multiple limits will be discussed in Sec. 192. The N and O limits have very long wave lengths, however, and can be observed only for the heaviest atoms. The ν/R values of some L and M limits are listed in Tables III(c), (d).

(c) *Fluorescent Spectra*.—Finally, the *fluorescent radiation* that Barkla observed to be emitted when X-rays fall upon material bodies is likewise easily explained in terms of the theoretical picture. A K electron may be ejected from the atom as the result of an absorption process as well as by cathode-ray bombardment. Such an act of absorption leaves a vacancy in the atom which may be filled subsequently by an electron dropping into it from the L shell, or from the M shell, and so on. Thus, as with cathode-ray excitation, characteristic radiation belonging to the entire K series of lines appears in fluorescence whenever any member of the series appears. The frequency of this fluorescent radiation should be lower than the frequency of the original incident radiation, in agreement with observation. Fluorescence is a comparatively unusual phenomenon in the visible region, but in the case of X-rays it is very common.

191. X-ray Energy Levels.—As in the spectroscopy of longer wave lengths, X-ray spectra are conveniently interpreted in terms of a set of *energy levels*. In the emission of a $K\alpha$ line, for example, an (ionized) atom undergoes a transition or “jumps” from a K energy

level to an L level, the difference in energy between these two levels being emitted as the $K\alpha$ photon. Although we often speak of an electron as making the jump, it is to be emphasized that the energy of each state or level involved in the transition really belongs to *the atom as a whole*. It would not be correct to think of this energy as belonging to a single electron, and it would be an unnecessary complication to endeavor to introduce a set of electronic energy levels.

When a K electron is removed from a neutral atom, the atom is raised into an atomic K level of much higher energy. The difference in energy between the K level and the neutral, normal level is equal to the work required to remove a K electron and leave it at rest at an infinite distance from the atom; thus it equals what we have called W_K or $h\nu_K$, where ν_K is the frequency of the K absorption limit. Similarly, there will be three slightly different L levels having energies greater than that of the normal level by amounts equal to the work of removing an L electron. Below the L levels, again, is a group of five M levels, associated with the removal of an M electron; and so on. Finally, just above the normal state lies the whole set of optical levels of the neutral atom, connected with the emission of the lines in the arc spectrum; and above this set of levels lies another set belonging, like the X-ray levels, to the ionized atom and associated with the *spark* spectrum in the visible and ultraviolet regions. Ordinary X-ray spectra constitute, in a sense, an extension of the spark spectrum to higher levels of energy, resulting from an electronic vacancy in the interior of the ionized atom rather than from the excitation of one of the outer electrons of the ion.

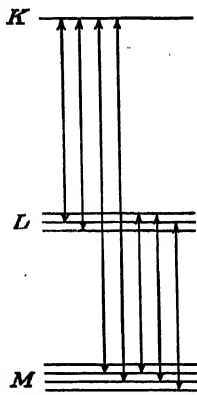


FIG. 159.—
Typical arrange-
ment of the higher
X-ray energy levels.

The typical distribution of atomic levels thus arrived at is shown diagrammatically in Fig. 159. Lines indicating transitions between X-ray levels on such diagrams are often drawn with an arrowhead at each end, to indicate that, in a sense, an electron moves upward on the diagram as the atom moves downward. Thus, in the emission of a $K\alpha$ line, the atom drops from its K level to one of its L levels, an electron simultaneously changing from an L shell to a K shell; for an $L\alpha$ line, an electron changes from M to L ; and so on. The $K\beta$ line results from the transition of the atom from the K to an M level; the loss of energy being larger, the $K\beta$ line is of higher frequency than the $K\alpha$ line. The $K\gamma$ line results from a transition, K to N . The transition from K to the normal state, representing capture of a free electron

of negligible kinetic energy into the vacancy in the K shell, would cause the emission of a frequency ν_K equal to that of the K absorption limit; this would represent the "series limit" for the K lines.

In a similar way, the lines of the L series are emitted by atoms in which a vacancy in the L shell has been caused by the impact of a cathode-ray electron or by the absorption of a photon. The atom then drops from its L energy level to the M level, or to some lower level, emitting a photon of radiation belonging to one of the L lines. Lines belonging to M, N, \dots series may similarly occur.

The energy-level diagram suggests at once a simple quantitative relation that should exist between *line frequencies* and *absorption limits*. It is obvious that the loss of energy by the atom when a $K\alpha$ line is emitted, *i.e.*, the $h\nu$ value of this line, is equal to the *difference* in the energies, or in the $h\nu$ values, of the K and L *absorption limits*. The same equality will exist for the frequencies, or for the values of ν/R . This is a conclusion that can easily be checked from the data. In columns 2 and 3 of Table IV are shown the wave lengths and the

TABLE IV.—RELATION OF THE FREQUENCIES OF THE $K\alpha_{1,2}$ LINES OF MOLYBDENUM TO THE FREQUENCIES OF THE K AND THE L ABSORPTION LIMITS

Absorption limits			$(\nu/R)_K - (\nu/R)_L$	Observed emission lines		
Limit	λ , X units	ν/R		Line	λ , X units	ν/R
K	618.48	1,473.4				
L_I	4,289.7	212.4	(1,261.0)	Missing	
L_{II}	4,712.0	193.4	1,280.0	$K\alpha_2$	712.10	1,279.7
L_{III}	4,904.2	185.8	1,287.6	$K\alpha_1$	707.83	1,287.4

values of ν/R for, respectively, the K, L_I, L_{II} , and L_{III} absorption limits of molybdenum [see Tables III(b),(c)]. In column 4 are shown the differences in ν/R values between the K limit and each of the three L limits. The second part of the table shows the directly observed values of λ and of ν/R for the $K\alpha_1$ and the $K\alpha_2$ lines of molybdenum. It is seen that the *observed value of ν/R for the $K\alpha_1$ line is almost exactly equal to the difference between the value of ν/R of the K limit and that of the L_{III} limit*. Similarly, the ν/R value of the $K\alpha_2$ line is equal to the difference between the value of ν/R of the K limit and that of the L_{II} limit. Putting this relation in symbols, we can write

$$\left(\frac{\nu}{R}\right)_{K\alpha_1} = \left(\frac{\nu}{R}\right)_K - \left(\frac{\nu}{R}\right)_{L_{III}},$$

$$\left(\frac{\nu}{R}\right)_{K\alpha_2} = \left(\frac{\nu}{R}\right)_K - \left(\frac{\nu}{R}\right)_{L_{II}};$$

or

$$\nu_{K\alpha_1} = \nu_K - \nu_{LIII}, \quad \nu_{K\alpha_2} = \nu_K - \nu_{LII}.$$

This example illustrates a law of great importance: *The frequencies of X-ray lines are given by the differences between the frequencies of absorption limits.* By means of this law it is possible to infer, from data on the lines, the relative positions of certain absorption limits which cannot be directly measured.

When absorption limits can be observed, on the other hand, they enable us, as we have seen, to locate the various energy levels relative to the normal state of the atom. Thus, from data on X-ray lines and absorption limits, the whole scheme of X-ray atomic levels can be built up, and this scheme then serves to represent a wide range of facts in a compact and clear form.

192. The Quantum Theory of X-ray Terms and Lines.—The simple picture of the atomic mechanism for the production of characteristic X-rays that was described above affords no explanation of the occurrence of several different L , M , . . . levels lying close together. This feature was not correctly understood until wave mechanics was developed. The basis for its explanation is furnished by the theory of atomic structure described in Secs. 127 to 129. As was there stated, however, the scope of this book is such that only a brief description of the theoretical results can be given here.

As explained in Sec. 128, to a first approximation the electrons in an atom can be supposed to be distributed, in accordance with the Pauli exclusion principle, among a number of representative 1-electron wave functions. Each of these functions can be chosen so as to be characterized by four quantum numbers (Sec. 153), $n l j m$. The first of these numbers, n , is always a positive integer; the values $n = 1, 2, \dots$ correspond to the various shells, which we have also denoted by K, L, M, \dots . The number l is 0 or a positive integer less than n ; it serves to designate a subshell. The number j can take on, for given $l > 0$, either of two values, $j = l + \frac{1}{2}$ or $j = l - \frac{1}{2}$; and m may have any one of the $2j + 1$ integrally spaced values from $m = j$ down to $m = -j$. For convenience, there is shown below the constitution

Shell	K	L		M				N			
n	1	2		3				4			
l	0	0	1	0	1	2	0	1	2	3	
j	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$ $\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$ $\frac{3}{2}$	$\frac{3}{2}$ $\frac{5}{2}$	$\frac{1}{2}$	$\frac{1}{2}$ $\frac{3}{2}$	$\frac{3}{2}$ $\frac{5}{2}$	$\frac{5}{2}$ $\frac{7}{2}$	
Electrons . . .	2	2	2 4	2	2 4	4 6	2	2 4	4 6	6 8	

of the first four shells when completely filled, the X-ray designation of each shell being given opposite "Shell," and the total number of electrons that have the same set of values of n , l , j being given opposite "Electrons."

Let us fix our attention now on a single subshell of a normal, neutral atom. This subshell is specified by a particular choice of n and l (e.g., $n = 3$, $l = 1$, the subshell containing 6 electrons). We wish to consider the effect of removing 1 electron.

When the subshell is full of electrons, there is only one possible choice that can be made of 1-electron wave functions to represent the electrons; it is necessary to use all of the wave functions that can be constructed for the given values of n and l . Under such circumstances the total angular momentum of all of the electrons in the subshell is easily shown to be zero; this is true likewise of the total orbital momentum of the electrons and of their spin momenta, and also for the sum of the angular momenta of both kinds.

If, now, 1 electron is removed from the subshell, there is a variety of states in which the incomplete subshell can be left, corresponding to the different combinations of the electronic wave functions that can be chosen for the remaining electrons. There are, in fact, just as many different possible states for the incomplete subshell as there are different types of electronic wave functions represented in the complete subshell, each state of the incomplete subshell corresponding to the omission of one of these functions. The resultant angular momenta, orbital and spin, of the electrons remaining in the incomplete subshell can then be added into a resultant, which we may indicate, according to spectroscopic terminology, as in Sec. 140, by the quantum number J . In addition, wave-mechanical theory indicates that the orbital angular momenta can be added separately into a resultant, with a quantum number L , and their spin momenta into a separate resultant characterized by a quantum number S (Secs. 141, 144). Corresponding to what was just said about the number of possible states for the subshell, however, the sets of values of J , L , and S that can occur for the remaining electrons in the subshell are exactly the same as they are for a single electron existing all by itself in the subshell. For this reason, it is customary to use, instead of the capital letters J , L , and S , the small letters j , l , and s , which would be used if only 1 electron were present. We can, in fact, think of the "hole" that is left by removing an electron as functioning like an electron, a conception that has far-reaching applications. The hole is sometimes thought of as formed by the removal of an electron

which already possessed the particular values of l and j in question. This idea must not be taken too literally, however; it does not imply, for instance, that jj coupling necessarily holds in the inner subshells of atoms.

The removal of an electron thus leaves the subshell in a state characterized by certain values of $n l j m$. Theory indicates, however, that the energy should not vary with m , so that for many purposes we may ignore this latter quantum number and retain only $n l j$. Furthermore, the energy, although varying greatly with the value of n , should vary only slightly with l and j . Thus, for an atom with a vacancy in a given shell (given n), there will exist a close group of energy levels whose number will be equal to the number of the possible sets of values of l and j . For $n = 1$ (K shell), the only values possible are $l = 0, j = \frac{1}{2}$; and there is only one level. For $n = 2$ (L shell), there is one level for $l = 0, j = \frac{1}{2}$, and also one each for $l = 1, j = \frac{1}{2}$ and $l = 1, j = \frac{3}{2}$, or three in all. For $n = 3$ (M shell), there are five possible levels, with $l = 0, j = \frac{1}{2}$; $l = 1, j = \frac{1}{2}$ or $\frac{3}{2}$; $l = 2, j = \frac{3}{2}$ or $\frac{5}{2}$; and so on. The number of the absorption limits as actually observed is thus explained.

Of course, the total number of levels that can occur for an actual atom is limited also by the number of electrons it contains. The uranium atom will exhibit the maximum number of levels. The energy-level diagram for ionized uranium, with energy plotted as ν/R on a logarithmic scale, is shown in Fig. 160. As one proceeds toward elements of lower atomic number, the electron shells of higher quantum number n progressively disappear (see Appendix III) and therewith, also, the corresponding energy levels, as well as the X-ray lines, originating from these levels. In copper, for example, there are no O levels and only one N level. One should not expect to find, therefore, in the X-ray spectrum of copper such lines (see Fig. 160 for their origin) as $L\beta_u, L\beta_7, L\gamma_6, L\gamma_8$, etc.

In addition to accounting for the general features of the energy-level diagram, wave mechanics makes definite predictions concerning the *relative probabilities of transitions* between the various levels. For a transition accompanied by what is called radiation of the dipole type, we have the same selection rules as for 1 electron (Sec. 153):

$$\Delta l = \pm 1, \quad \Delta j = 0 \text{ or } \pm 1.$$

Experimentally, it is observed that between many pairs of levels transitions do not occur. For example, in Table IV, the emission line corresponding to the transition $K \rightarrow L_I$ is listed as "missing." In

endeavoring to assign values of l and j in such a way as to bring the theoretical predictions into harmony with the facts.

Referring to the uranium diagram in Fig. 160, which is typical of that for all atoms except for the absence of certain outer levels as explained above, we note that transitions occur between the K level, with $l = 0$, and the L_{II} and L_{III} levels, giving rise to the $K\alpha_1$ and $K\alpha_2$ lines. Hence, it must be that $l = 1$ for L_{II} and L_{III} , in order to make $\Delta l = 1$. For L_I , there remains then only $l = 0$, as indicated at the left side of the figure. The $K\beta$ and $K\gamma$ "lines," also, are actually doublets like $K\alpha$, which can be resolved in heavy atoms; they fix $l = 1$, likewise, for M_{II} , M_{III} , N_{II} , N_{III} . The remaining M levels, M_I , M_{IV} , and M_V , must then consist, in some order, of $(0, \frac{1}{2})$, $(2, \frac{3}{2})$ and $(2, \frac{5}{2})$ (i.e., of levels with $l = 0$ and $j = \frac{1}{2}$, etc.). These latter three levels could all combine with L_{II} , L_{III} , i.e., with $(1, \frac{1}{2})$ and $(1, \frac{3}{2})$, without violating the rule that $\Delta l = \pm 1$. The rule that $\Delta j = 0$ or ± 1 , however, limits $(2, \frac{5}{2})$ to combining with $(1, \frac{3}{2})$ only. It is found experimentally that M_V combines only with L_{III} , giving rise to the brightest of the L lines, called $L\alpha_1$, whereas M_I and M_{IV} combine with both L_{II} and L_{III} (lines $L\eta$, $L\zeta$, $L\beta_1$, $L\alpha_2$, the first two very weak). Hence, it must be that $M_V = (2, \frac{5}{2})$, and also that $L_{II} = (1, \frac{1}{2})$, so that, between L_{II} and M_V , $\Delta j = 2$ and the transition is forbidden. Thus we have identified L_I and L_{II} ; L_{III} is then also identified, by a process of elimination, as $(1, \frac{3}{2})$.

If enough lines of the M group were known, we could continue in this manner and identify all of the M levels. In default of such knowledge, we may fall back on the conclusion drawn from wave-mechanical perturbation theory that the levels should lie in the same order in all shells, and with levels of larger j at lower energies. This principle agrees with all that can be learned from the observed lines; and it suffices to complete the arrangement of the levels, as shown in Fig. 160.

The values of n , l , and j thus associated with the various-X-ray levels are shown at the left in Fig. 160. It may be remarked that the corresponding spectroscopic notation for atomic levels is also sometimes employed, the correspondence between this notation and the older X-ray symbols being as follows (Sec. 142; all the terms here are doublet terms):

K $1S_{\frac{1}{2}}$	L_I $2S_{\frac{1}{2}}$	$L_{II,III}$ $2P_{\frac{1}{2},\frac{3}{2}}$	M_I $3S_{\frac{1}{2}}$	$M_{II,III}$ $3P_{\frac{1}{2},\frac{3}{2}}$	$M_{IV,V}$ $3D_{\frac{3}{2},\frac{5}{2}}$
N_I $4S_{\frac{1}{2}}$	$N_{II,III}$ $4P_{\frac{1}{2},\frac{3}{2}}$	$N_{IV,V}$ $4D_{\frac{3}{2},\frac{5}{2}}$	$N_{VI,VII}$ $4F_{\frac{5}{2},\frac{7}{2}}$		

Certain predictions as to the *relative intensity* of X-ray lines may also be derived from the theory. Lines in which l and j change in the same sense should be stronger than those in which l and j change in opposite directions; and the lines for the largest values of l and j should be, as a rule, the strongest. These conclusions from the theory agree, in a general way, with the data. The order of the subscripts commonly used in designating lines, $\alpha, \beta, \gamma, \delta, \dots$, indicates, in general, decreasing intensity; lines such as η and l are very weak. In the case of lines that start from a common initial level, furthermore, the theory leads to *quantitative* predictions concerning the relative intensities. When the initial level is the same for two lines, uncertainties in regard to the relative probabilities of excitation do not arise. Thus, theoretically, $K\alpha_1$ should be twice as strong as $K\alpha_2$; and experimentally this is very nearly true.

Experimental work on the relative intensities of X-ray lines is, however, usually very difficult, because of the numerous troublesome corrections that have to be made to the observed intensities.¹

It should be remarked in conclusion that the dipole selection rules, employed above, do not hold in all cases, in agreement with theoretical predictions. Many weak lines "forbidden" by these rules have been observed; especially in the spectra of the heavier elements, lines have been reported² corresponding to transitions $L_{III} \rightarrow N_{II}$ and $L_{III} \rightarrow N_{III}$, for each of which $\Delta l = 0$. Such lines are called "quadrupole" lines. They are subject to a different set of selection rules (e.g., $\Delta l = 0$ or ± 2).

193. The Continuous X-ray Spectrum.—The characteristic line spectrum emitted by an X-ray target as observed is superposed upon a *continuous spectrum*. This appears clearly, for example, in Fig. 155. The positions, i.e., wave lengths, of the *lines* are determined solely by the material of the target; their *intensity* is determined, for a given target material and tube current, by the voltage applied to the tube. On the other hand, the *wave-length* characteristics of the *continuous spectrum* are quite independent of the *material* of the target but are determined by the voltage applied to the tube. The *intensity* of the continuous spectrum, for a given tube current, is dependent both on the target material and on the applied voltage, as well as on the thickness of the target. In this section, we shall give a brief survey of some of the more important facts concerning the continuous spectrum.

¹ PARRATT, *Phys. Rev.*, vol. 54, p. 99 (1938); see also Sec. 194.

² IDEI, *Tôhoku Univ., Sci. & Technol. Reports*, vol. 19, p. 559 (1930); KAUFMAN, *Phys. Rev.*, vol. 40, p. 116 (1932).

A series of four spectral-energy distribution curves, recorded with an ionization spectrometer for the radiation from a tungsten target, and for applied voltages of 20,000, 30,000, 40,000, and 50,000 volts, respectively, is shown in Fig. 161.¹ The spectral region of this figure lies between the *K* and *L* series lines of tungsten. Starting at the long-wave-length side, the curves rise to a maximum and then drop rapidly

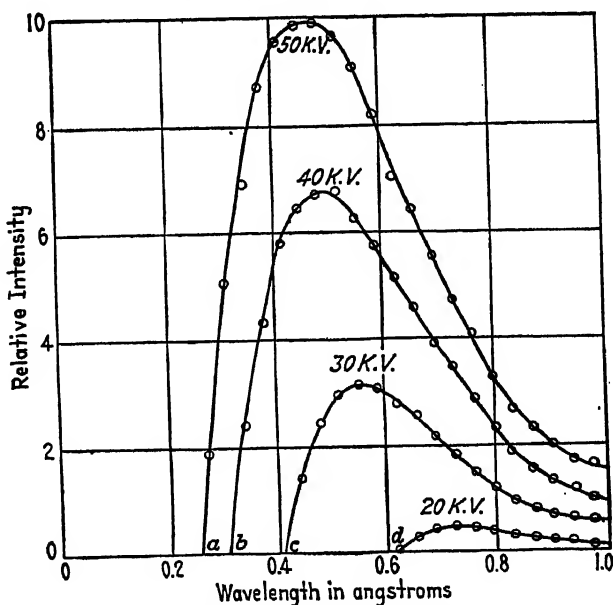


FIG. 161.—Ulrey's curves for the distribution of energy in the continuous X-ray spectrum of tungsten at various voltages.

toward zero; the position of the maximum depends on the applied voltage. The curves are seen to meet the axis at finite angles, as is shown at the intersections *a*, *b*, *c*, and *d*, respectively. These intersections can be very accurately measured by making readings nearer the axis² than the readings shown in Fig. 161. It is seen that the intersections come at shorter wave lengths, the higher the voltage.

A very simple relation exists between these limiting frequencies and the applied voltage. This relation is shown graphically in Fig. 162, in which limiting frequencies are plotted as ordinates against applied voltages as abscissas. The graph is a straight line passing

¹ These curves are from the measurements of Ulrey, *Phys. Rev.*, vol. 11, p. 401 (1918). They are subject to certain corrections to be mentioned later.

² See DUANE and HUNT, *Phys. Rev.*, vol. 6, p. 166 (1915).

through the origin; the limiting frequency is strictly proportional to the applied voltage, the empirical equation of the curve being

$$\nu_0 = 2.43 \times 10^{14} V', \quad (271)$$

where ν_0 is the *maximum* frequency which an applied voltage V' (in volts) can generate.

The existence of this sharp "cutoff," at a limiting minimum wave length, is hard to explain on the basis of classical theory. The continuous spectrum is believed to be emitted as a consequence of the deflection of the cathode electrons by the strong fields surrounding

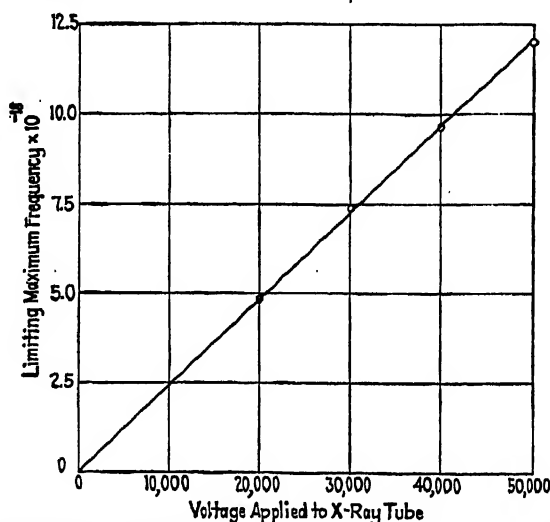


FIG. 162.—The relation between the limiting frequency and the applied voltage.

the nuclei of the atoms of the target, in contrast to the line spectrum, which is emitted by electrons of the target itself in atoms which have been ionized. According to classical theory, the acceleration of an electron as it undergoes deflection in the field of a nucleus will cause the emission of a pulse of radiation. The energy-distribution curve for such a pulse can be found by making a Fourier analysis of the pulse into monochromatic wave trains. Now a form of pulse can be invented which produces a sharp cutoff of the distribution curve;¹ but no good reason can be assigned to explain why just this form of pulse should always be emitted in consequence of the deflection or stoppage of a cathode particle. Quantum theory, on the other hand, easily explains the existence of the sharp cutoff.

¹ KENNARD, *Phys. Zeits.*, vol. 24, p. 372 (1923).

According to quantum theory, our conception of the process by which the negative acceleration of cathode electrons produces radiation must be modified in somewhat the same way as the classical theory of the emission of a spectral line by an atom. According to classical theory, a spectral line represents radiation that is emitted continuously, with a frequency equal to the frequency of vibration of an electron, the energy of which gradually diminishes during the emission of the radiation. According to quantum theory, on the other hand, the radiation is emitted as a quantum or photon of energy $h\nu$, and its frequency ν is determined by the Einstein condition that

$$h\nu = \Delta W,$$

where ΔW is the total amount of energy lost by the electron. Thus a cathode electron passing through the strong electric field near the nucleus of an atom in an X-ray target is not to be thought of as undergoing continuous acceleration, accompanied by the emission of radiation, as is required by classical theory. The electron may actually pass through the field undeviated in direction, but it is much more likely to issue from the atom in a different direction from its direction of approach. It may suffer no change in its energy; and in this case no radiation is emitted. Or, it may suffer a decrease ΔW in energy; in this latter case the energy lost is emitted as a photon of frequency ν given by Einstein's equation, $h\nu = \Delta W$. The amount of energy lost cannot exceed the total kinetic energy with which the electron enters the target, but it may have any value smaller than this.

It follows that the *maximum frequency* ν_0 of the X-rays will be the frequency of a photon emitted when the electron is brought entirely to rest as the result of a single elementary process. For the energy with which the electron approaches the atom, we can write eV , e being the electronic charge and V the potential difference through which it falls in passing from the cathode of the tube to the target. The maximum frequency ν_0 present in the X-rays will thus be determined by the equation

$$h\nu_0 = eV. \quad (272)$$

In addition to this maximum frequency ν_0 , we should then expect also a whole spectrum of *lower* frequencies, emitted by electrons which lose only a part of their energy in a single encounter with a nucleus, in the form of a smaller quantum $h\nu$. A further reason for the emission of frequencies lower than ν_0 lies in the fact that, in the vast majority of cases, an electron will experience many collisions with atoms of the target before being brought to rest. At each of these collisions some of the energy eV is dissipated. Thus, a great deal of radiation is

emitted by electrons which impinge upon atoms with incident energy less than eV . If the target is sufficiently thin, however—say of very thin gold foil—only a few of the electrons of the incident cathode stream will collide with atoms in it, most of them passing through the target undeviated. Thus slowly moving electrons will not be present in a *thin* target to the same degree as in a *thick* one. Accordingly, we should expect that a greater proportion of the energy in the continuous spectrum from thin targets should lie near the ν_0 limit than from thick targets. This is in agreement with experiment.

In the continuous spectrum from a *very thin* target, as a matter of fact, experiment indicates, in agreement with the wave-mechanical computation of Sommerfeld, that the maximum of the energy-distribution curve occurs *at the limiting wave-length*. ν_0 *itself*. On the short-wave side of λ_0 , the curve drops abruptly to the axis of abscissas, whereas toward longer waves it falls nearly in proportion to $1/\lambda^2$, as illustrated in Fig. 163. The curves for a thick target, as in Fig. 161, can be regarded as arising from the superposition of many elementary curves, such as that in Fig. 163, with various values of λ_0 .^{*} In such a target it might be expected that there would be a *most probable* type of collision which would correspond to the peak or maximum of the energy-distribution curve.

With the *shape* of the energy-distribution curve we are not, for the moment, concerned. Rather, let us return to the empirical equation (271) and to the curve (Fig. 162) which showed, in complete agreement with Eq. (272), that $V \propto \nu_0$. Writing Eq. (272) in the form

$$\nu_0 = \frac{e}{h} V = \frac{e}{h} \frac{V'}{300},$$

where V is in electrostatic units and V' is in volts, we see that the numerical constant in Eq. (271) should be equal to the multiplier of V' in this last equation. That is, we should have

$$\frac{e}{300h} = 2.43 \times 10^{14},$$

from which, if we use the old value, $e = 4.77 \times 10^{-10}$ e.s.u.,

$$h = 6.55 \times 10^{-27},$$

^{*} Cf. COMPTON and ALLISON, *op. cit.*, p. 90.

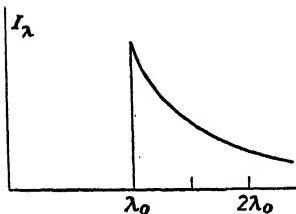


FIG. 163.—Approximate distribution of energy in the continuous X-ray spectrum from a very thin target, near the short-wave-length limit.

or, if we use the modern value, $e = 4.80 \times 10^{-10}$ e.s.u.,

$$h = 6.58 \times 10^{-27}.$$

At the present time (1940), the best value of h as derived from precise measurements of the limiting X-ray frequency and a modern value of e of about 4.803×10^{-10} , is¹ in good agreement with values obtained in other ways.

194. Intensity of the Continuous Spectrum.—Determinations of the total intensity, and of the distribution of intensity in the continuous X-ray spectrum are by no means so satisfactory as the measurements of minimum wave lengths. This is due to the fact that numerous troublesome corrections have to be made to the *observed* energy-distribution curves, such as those shown in Fig. 161, before the true curve can be obtained.² Among these necessary corrections may be mentioned the following:

1. *The variation with wave length of the absorption of the radiation in its path from the atom of the target in which the radiation originates to the ionization chamber* where the intensity of the radiation is measured. This path includes the material of the target itself, since the X-rays originate at finite (though very small) depths below the surface, the walls of the X-ray tube and the window of the ionization chamber, and the air between the tube and the chamber.

2. *The reflecting power of the crystal grating*, which varies with wave length.

3. *Incomplete absorption in the ionization chamber.* Either an ionization chamber of sufficient length to absorb the entire X-ray beam must be used, or else corrections must be made for variation with wave length of the absorption within the chamber, since the *observed* ionization current is proportional to the energy *absorbed* in the chamber.

4. *"Second-order reflection."* According to Eq. (268), which may take the form

$$\lambda = \frac{2d \sin \theta}{n},$$

when the crystal in the spectrometer [Fig. 154 (a)] is set at a given angle θ , a *series* of wave lengths λ will be reflected for different values 1, 2, 3, . . . of n , *provided* those several wave lengths are present in the incident radiation. Suppose, for example, that when the voltage applied to the X-ray tube is 50 kilovolts, the crystal is set at

¹ Cf. Dunnington, *Rev. Modern Physics*, vol. 11, p. 65 (1939), and for the best experimental work, DuMond and Bollman, *Phys. Rev.*, vol. 51, p. 400 (1937).

² These corrections, however, in no wise affect the determination of minimum wave lengths.

such an angle $\theta_{0.6}$ as to give a *first-order* reflection (*i.e.*, $n = 1$) of $\lambda = 0.6$ angstrom. According to Fig. 161, the minimum wave length present in the X-ray spectrum generated by 50 kilovolts is 0.250 Å. Under these conditions, there will be reflected into the ionization chamber the *second-order* wave length $\lambda = 0.3$ Å, in addition to the first-order wave length $\lambda = 0.6$ Å. The *observed intensity* when the spectrometer is set at $\lambda = 0.6$ Å is, therefore, the sum of the intensities due to both wave lengths. For an applied voltage of 30 kilovolts, however, only first-order reflection is present at $\theta_{0.6}$, and no correction is necessary.

5. *Secondary production of X-rays.* If one is interested primarily in the radiations produced by impacts of the incident cathode rays alone, corrections must be made for the radiations arising from impacts of the various types of secondary electrons produced in the target itself, *e.g.*, electrons ejected from the atoms of the target by the impact of cathode rays, or by photoelectric ejection, or by ejection as Compton electrons in connection with the scattering of the X-rays in the target [Sec. 198(a)].

6. *Characteristic lines.* There generally exist several characteristic X-ray lines superimposed on the continuous spectrum; the intensities of such lines must be subtracted from the intensity of the beam as observed.

Without discussing the methods of making these several corrections,¹ the final results may be summarized as follows:

- a. The total radiation, *i.e.*, the area under the (corrected) energy-distribution curves, is, for a given target material and tube current, very nearly proportional to the square of the applied voltage.
- b. At a given voltage and tube current, the total radiation varies roughly as the first power of the atomic number of the target.
- c. The wave length λ_m at which a given energy-distribution curve for a thick target attains its maximum is roughly 1.5 times the wave length at which the same curve meets the wave-length axis.
- d. Wagner and Kulenkampff² give the following empirical formula for the intensity I_ν of the X-ray energy at frequency ν from a target of atomic number Z when bombarded by electrons which have fallen through a potential V :

$$I_\nu = CZ(\nu_0 - \nu) + BZ^2,$$

¹ See KIRKPATRICK, *Phys. Rev.*, vol. 22, p. 414 (1923).

² *Ann. d. Physik*, vol. 68, p. 369; vol. 69, p. 548 (1922).

where ν_0 is the maximum frequency determined by the quantum relation $eV = h\nu_0$, and C and B are constants independent of voltage and target material.

The efficiency of an X-ray tube (with thick target) is very low. If we define the efficiency as the fraction of the energy carried by the cathode electrons which is converted into X-ray energy, it varies from about 0.01 to 0.1 percent.

INTERACTIONS OF X-RAYS WITH ATOMS

195. The Absorption of X-rays.—In contrast with the apparently chaotic state of affairs in regard to the absorption of light in the visible or near-visible portions of the spectrum, we find comparative simplicity in the empirical laws of the absorption of X-rays. Measurements of the absorption coefficient for a monochromatic beam of parallel rays are readily made by use of the ionization spectrometer [Fig. 154(a)]. For a given crystal angle θ and, therefore, wave length λ , the ionization current is measured both with and without a sheet of absorbing material of known density ρ and thickness x placed in the path of the beam, say between the two slits S_1 and S_2 . These measurements give, respectively, I and I_0 in the equation [cf. Eq. (266) in Sec. 185]

$$I = I_0 e^{-\mu x} = I_0 e^{-\frac{\mu}{\rho} \rho x}, \quad (273)$$

from which the linear absorption coefficient μ or the mass absorption coefficient μ/ρ may be computed. In this way, being careful to use such a tube voltage as to eliminate second order reflections, one may obtain values of μ/ρ for various wave lengths and various substances. The mass absorption coefficient μ/ρ is much more commonly used than the linear coefficient.

(a) *Properties of μ/ρ .*—Figure 164 shows roughly the mass absorption coefficient of lead in the wave-length range $0.1 < \lambda < 1.2$ Å. Beginning at point o , μ/ρ rises rapidly with increasing wave length until point a , corresponding to $\lambda = 0.1405$ Å and $\mu/\rho = 8$ (about) is reached, at which the value of μ/ρ suddenly drops to point a' . This is the K absorption limit (Secs. 185, 190). Up to this point, absorption accompanied by ejection of K electrons has constituted the major part of the absorption. At longer wave lengths, this K absorption does not occur at all, hence the drop in μ/ρ .

With further increase in wave length, however, the absorption again increases rapidly, being mostly due now to ionization of the atoms in the L shell, until at point b , corresponding to $\lambda = 0.780$ Å,

the L_I absorption limit is reached, at which there occurs another drop in the absorption to point b' . From b' to c , absorption still occurs which is accompanied by ionization leaving atoms in L_{II} or L_{III} states; *i.e.*, the atoms in question are raised to their L_{II} or L_{III} energy levels. None, however, is raised to the L_I level. Similar drops or "breaks" occur at cc' ($\lambda = 0.813$ Å) and at dd' ($\lambda = 0.950$ Å), the L_{II} and L_{III} limits, respectively; from c' to d , absorption into the

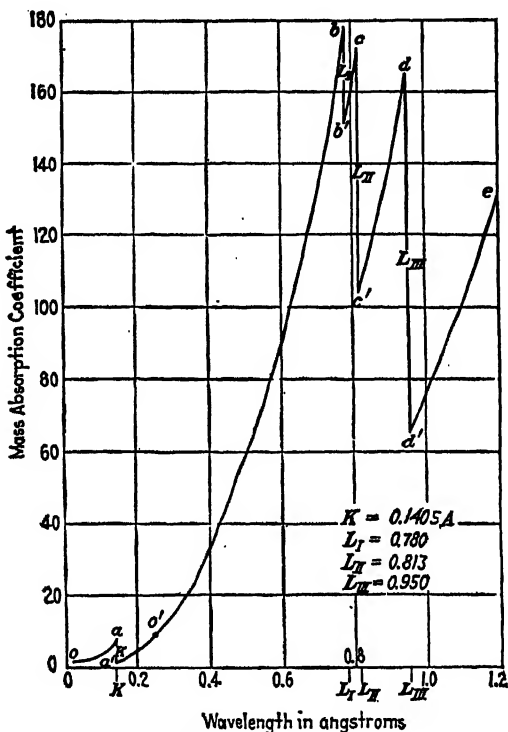


FIG. 164.—K and L absorption limits of lead.

L_{III} state is still occurring, but beyond d' absorption with ionization in the L shell ceases altogether. Beyond point d' , however, the absorption again increases rapidly. If we could follow, by direct measurement, the absorption beyond point e , we should find that, in the region $3.2 < \lambda < 5.0$ Å, a group of five "breaks" occurs, representing the five M absorption limits; beginning at about 14 Å, there would come the group of seven N limits; etc.

Curves similar to Fig. 164 are obtained for the absorption of X-rays in other substances, the respective discontinuities or limits occurring at longer wave lengths, the lower the atomic number of the absorber.

In fact, a curve plotted between the square root of the frequency of a given limit and atomic number is nearly a straight line similar to Moseley's curve (Fig. 156) for line spectra. Such a curve for the K limit, from Mg ($Z = 12$) to Th ($Z = 90$) is shown in¹ Fig. 165, in which, as ordinates, are plotted $\sqrt{\nu/R}$, where R is the Rydberg constant. The graph is seen to be nearly, but not quite, a straight line. For comparison, there are also shown the corresponding graphs for the L_I , L_{II} , and L_{III} limits from Ag ($Z = 47$) to Th ($Z = 90$).

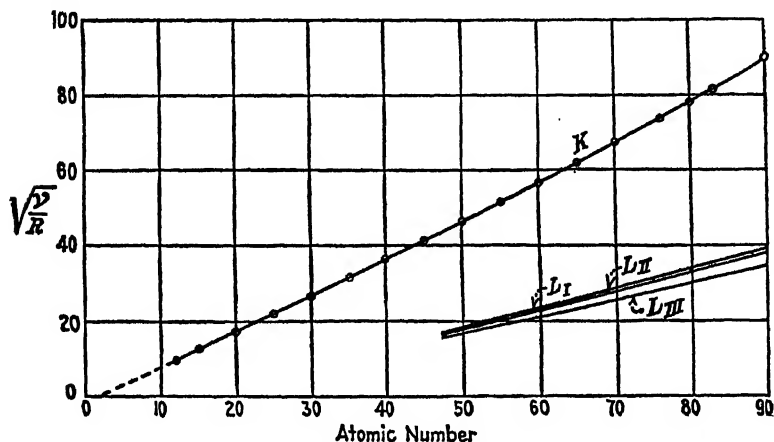


FIG. 165.—The Moseley diagram for the K and L absorption limits.

Not only is there regularity in the wave lengths of the absorption limits from one element to another, but there is also a remarkable regularity in the *magnitude* of the mass absorption coefficient, from wave length to wave length in a given element, and from one element to another. Figure 166 shows, on a much larger scale, the mass absorption coefficients of lead in² the region o to o' of Fig. 164, plotted as a function of the *cube* of the wave length. The two parts of the graph on each side of the K absorption limit aa' are seen to be *straight lines*, with very nearly equal intercepts of about $\mu/\rho = 1$. We may, accordingly, write the empirical equation for the variation of μ/ρ with λ in the regions oa and $a'o'$, respectively, as

$$\frac{\mu}{\rho} = k\lambda^3 + b, \quad \frac{\mu}{\rho} = k'\lambda^3 + b, \quad (274a,b)$$

where k and k' are the slopes of the respective lines and b is a constant.

¹ Data from SIEGBAHN, *loc. cit.*

² RICHTMYER, *Phys. Rev.*, vol. 27, p. 1 (1925).

To a first approximation,¹ the mass absorption coefficient for a given absorber is a linear function of the cube of the wave length, due regard being given to the spectral region in which the equations apply.

A quantity of more fundamental significance than μ or μ/ρ is the *atomic absorption coefficient* μ_a . This quantity represents the absorption due to a single atom and is obtained by dividing the mass absorp-

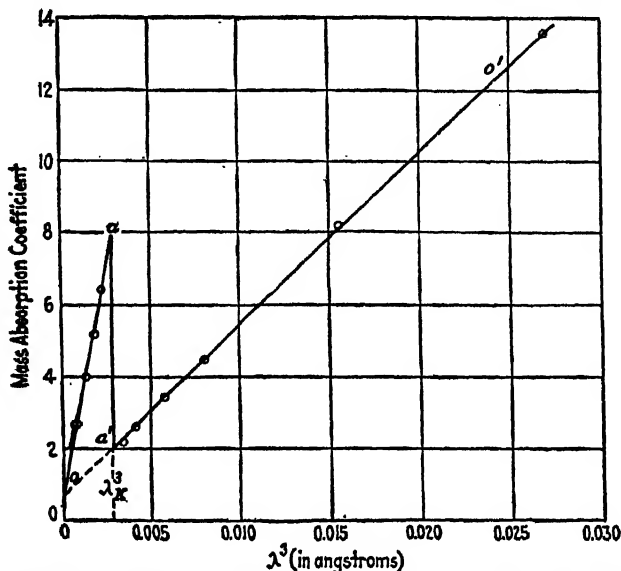


FIG. 166.—Mass absorption coefficients of lead as a function of wave length, showing the K discontinuity and the linear relation between μ/ρ and λ^3 .

tion coefficient μ/ρ by the number of atoms per gram, N_0/A , where N_0 is Avogadro's number and A is the atomic weight:

$$\mu_a = \frac{\mu/\rho}{N_0/A}.$$

Equations (274a,b) may then be written

$$\mu_a = k_a \lambda^3 + b_a, \quad \mu_a = k'_a \lambda^3 + b_a. \quad (275a,b)$$

(More commonly the symbol σ_a is written for b_a , for the reason stated below.) In terms of μ_a a second type of regularity has been found to exist in the absorption coefficient. The constants k_a and k'_a vary approximately as the fourth power of the atomic number.² Accord-

¹ There are slight systematic departures from the λ^3 law. See COMPTON and ALLISON, *op. cit.*; CUYKENDALL, *Phys. Rev.*, vol. 50, p. 105 (1936); JONES, *Phys. Rev.*, vol. 50, p. 110 (1936); ANDREWS, *Phys. Rev.*, vol. 54, p. 944 (1938).

² BRAGG and PIERCE, *Phil. Mag.*, vol. 28, p. 626 (1914); RICHTMYER and WARBURTON, *Phys. Rev.*, vol. 22, p. 539 (1923).

ingly, Eq. (275a,b) may be written in the following form, λ_K and λ_L denoting, respectively, the K and L absorption limits:

$$\lambda < \lambda_K: \quad \mu_a = C_a Z^4 \lambda^3 + b_a \quad (276a)$$

$$\lambda_K < \lambda < \lambda_L: \quad \mu_a = C'_a Z^4 \lambda^3 + b'_a \quad (276b)$$

Here, when λ is in centimeters, $C_a = 2.25 \times 10^{-2}$, $C'_a = 0.33 \times 10^{-2}$.

It seems surprising that such simple relations should hold, even approximately, for the absorption coefficients. The theoretical treatment of this topic is not in a very satisfactory state.¹

(b) *Physical Interpretation of the Absorption Formulas.*—As we have seen, the so-called “absorption” of X-rays is due partly to the photoelectric effect, to be described in the next section, and partly to scattering (Secs. 184, 185). Hence, we may conveniently write

$$\mu = \tau + \sigma, \quad \mu_a = \tau_a + \sigma_a,$$

where τ or τ_a represents the part of the absorption that is due to the photoelectric effect and σ or σ_a the part that is due to scattering. According to classical theory the scattering should be independent of frequency [Sec. 184(a)], and for this reason σ or σ_a has often been identified with the constant term b or b_a in the empirical formulas (275a,b) or (276a,b). The term in λ^3 is then identified with the photoelectric absorption. We shall see, however, that in reality the scattering coefficient varies considerably with wave length, so that such an interpretation of the empirical formulas can possess only approximate validity.

196. The Photoelectric Effect for X-rays.—If part of the absorption of X-rays is due, as we have assumed, to a photoelectric effect, it should be possible to observe the photoelectrons thus produced, and Einstein's photoelectric equation should be applicable to them. In the ordinary photoelectric effect, the electrons come from among the “free” electrons in the metal (Sec. 57), and their maximum kinetic energy, when produced by a beam of light of frequency ν , is $\frac{1}{2} m v_m^2$ as given by Eq. (53) in Sec. 49:

$$\frac{1}{2} m v_m^2 = h\nu - \omega_0,$$

where ω_0 is the work required to remove an electron from just under the surface of the metal (h = Planck's constant). When photoelectrons are produced by X-rays, on the contrary, they will come (at least principally) from the deeper-lying shells of the atoms. If a photoelectron is ejected from the K shell by X-rays of frequency ν , the amount of energy absorbed from the radiation is $h\nu$; but work equal

¹ See COMPTON and ALLISON, *op. cit.*

to $h\nu_K$ must be done by the electron in escaping from the atom, ν_K being the frequency of the K absorption limit. Hence the electron will emerge from the atom with kinetic energy equal to $h\nu - h\nu_K$. If the atom lies on the surface of the absorbing material, the electron may escape into the surrounding space with this amount of energy (more exactly with energy $h\nu - h\nu_K - \omega_0$, but ω_0 is usually negligibly small). Otherwise it may lose part of its energy in passing through a layer of matter. Thus the *maximum* kinetic energy with which photoelectrons produced by absorption in the K shell may emerge from the absorbing material will be

$$\frac{1}{2}mv_m^2 = h\nu - h\nu_K = h(\nu - \nu_K) \quad (277)$$

(provided ω_0 may be neglected). If $\nu < \nu_K$, no photoelectrons at all can be ejected from the K shell. As ν increases above ν_K , the maximum energy of the photoelectrons increases linearly, just as in the ordinary photoelectric effect.

The *number* of photoelectrons ejected from the K shell of atoms of a given absorber, however, decreases as ν increases, thereby causing the known decrease in the absorption coefficient as ν recedes from ν_K . In all cases this number is strictly proportional to the intensity of the X-rays.

The photoelectrons ejected from the L shell consist of three slightly different groups, with maximum energies corresponding to the three L absorption limits, ν_{LI} , ν_{LII} , ν_{LIII} ; the maximum kinetic energies in the three groups are

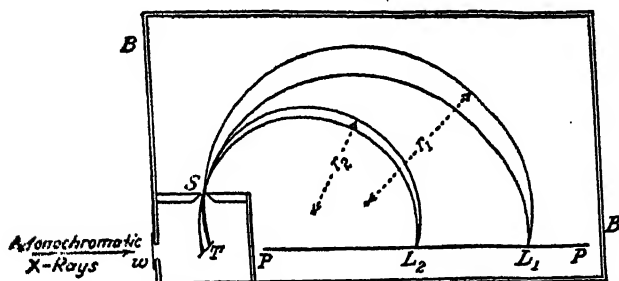
$$\begin{aligned} \frac{1}{2}mv_{mI}^2 &= h(\nu - \nu_{LI}), & \frac{1}{2}mv_{mII}^2 &= h(\nu - \nu_{LII}), \\ \frac{1}{2}mv_{mIII}^2 &= h(\nu - \nu_{LIII}). \end{aligned} \quad (278a,b,c)$$

The differences in these maximum energies arise from differences in the quantum state in which the remaining L electrons are left upon the departure of the ejected one. Similarly, from the M shell there are five groups; and so on. If we start with X-rays of very high frequency, photoelectrons of all kinds are produced. If the frequency is then decreased, as it passes ν_K , the K photoelectrons disappear; as it passes the L limits, the three groups of L photoelectrons disappear in turn; and so on.

Among the experiments demonstrating these facts may be mentioned those of Robinson and his collaborators,¹ whose apparatus is

¹ ROBINSON and RAWLINSON, *Phil. Mag.*, vol. 28, p. 277 (1914); ROBINSON, *Roy. Soc., Proc.*, vol. 104, p. 455 (1923); *Phil. Mag.*, vol. 50, p. 241 (1925); ROBINSON and CASSIE, *Roy. Soc., Proc.*, vol. 113, p. 282 (1927); ROBINSON and YOUNG, *Roy. Soc., Proc.*, vol. 128, p. 92 (1930).

Diagrammatically in Fig. 167. A beam of X-rays of frequency¹ through a thin window W a highly evacuated brass box BB upon a target T of the material under investigation. Photo-



Robinson's magnetic spectrograph for studying the photoelectric action of X-rays.

are expelled from the surface of T in all directions and with velocities, as indicated by Eqs. (277), (278a,b,c). The whole apparatus is placed in a known magnetic field at right angles to the plane of the paper, and can be varied at will. The photoelectrons all describe circles in the field. Some of them pass through the narrow slit S and finally strike the photographic plate PP . Electrons leaving T have velocities v , they will move in circles of radii r , and will strike the plate at points L_1 and L_2 . As shown in the figure, the arrangement is such as to "focus" electrons from the different parts of the target with the same velocity onto the plate at such positions as L_2 , the diameter of the circle being the distance between S and L_1 or L_2 . In an experiment such as this, the velocity of electrons is so great that variation of mass with velocity must be taken into account. This is easily done as follows. As the electron

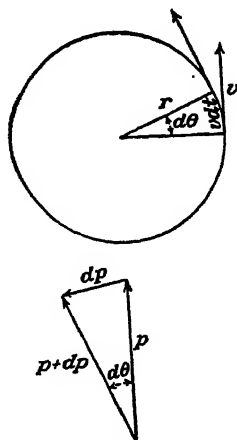


FIG. 168.

is extremely difficult by present experimental means to get a strictly monochromatic beam of X-rays of sufficient intensity for such an experiment as this. Usually, use is made of the fact that, with suitable exciting voltage, the $K\alpha$ lines from an X-ray target are much more intense than the accompanying radiation of other wave lengths and, therefore, serve as a "monochromatic" beam, especially if the beam is first passed through a filter of suitable thickness, the length of whose K limit is just shorter than the wave length of the $K\alpha$ lines. (i.e., *Phys. Rev.*, vol. 10, p. 661 (1917).]

goes around in a circle under the influence of the magnetic field, its momentum p is constant in magnitude but continually changes in direction. We can imagine that the electron receives continual vector increments of momentum directed toward the center of the circle. As the direction of motion changes by $d\theta$, an increment dp of momentum must be added of magnitude (Fig. 168)

$$dp = p d\theta.$$

The force on the electron is evH , e being the electronic charge in electromagnetic units and v its velocity. Hence during a time dt the electron receives momentum

$$dp = evH dt.$$

Thus

$$dp = p d\theta = evH dt.$$

But (Fig. 168)

$$d\theta = \frac{v dt}{r}.$$

$$\therefore p = erH. \quad (279)$$

Now according to the relativistic formula [Eq. (67) in Sec. 67]

$$p = \frac{mv}{(1 - v^2/c^2)^{1/2}}, \quad (280)$$

where m stands for the rest mass of the electron (0.911×10^{-27} gram); and Eq. (68) in Sec. 68 gives for its kinetic energy

$$T = mc^2 \left[\frac{1}{(1 - v^2/c^2)^{1/2}} - 1 \right].$$

From (280)

$$\frac{p^2}{m^2 c^2} = \frac{v^2/c^2}{1 - v^2/c^2}, \quad \therefore \frac{p^2}{m^2 c^2} + 1 = \frac{1}{1 - v^2/c^2}.$$

Hence we can also write¹

$$T = mc^2 \left[\left(1 + \frac{p^2}{m^2 c^2} \right)^{1/2} - 1 \right]; \quad (281)$$

and for an electron revolving in a magnetic field, by (279),

$$T = mc^2 \left[\left(1 + \frac{e^2 r^2 H^2}{m^2 c^2} \right)^{1/2} - 1 \right]. \quad (282)$$

(Here e is in electromagnetic units.)

¹ Calculations from such formulas are most easily made thus:

$$T = mc^2 [(\cos \tan^{-1} p/mc)^{-1} - 1].$$

By means of this last formula, values of the kinetic energy T of the photoelectrons can be calculated from measured values of r , for comparison with $h(\nu - \nu_A)$, where ν_A is any absorption limit. If losses of energy by the electrons within the absorber are negligible, we can substitute values of T so found in the photoelectric equation

$$T = h(\nu - \nu_A)$$

and, knowing ν , can obtain values of the absorption limit ν_A , for comparison with values measured spectroscopically.

TABLE V.—SOME X-RAY ABSORPTION LIMITS DETERMINED FROM THE MAGNETIC SPECTRUM OF PHOTOELECTRONS (AFTER ROBINSON); COMPARISON WITH SPECTROSCOPIC METHOD

Level	Uranium Z = 92		Tungsten Z = 74		Silver Z = 47		Copper Z = 29	
	Robin- son	Spec.	Robin- son	Spec.	Robin- son	Spec.	Robin- son	Spec.
K	660.1	661.6
L _I	891.3	890.8	279.4	282.0	79.9	81.0
L _{II}	850.2	849.9	258.7	261.3	68.3	70.3
L _{III}	750.8	751.3	245.8	248.6		68.9
M _I	408.7	408.5	208.1	207.3	47.1	54.7	5.8	8.9
M _{II}	380.1	381.5	189.1	189.3		46.1		5.7
M _{III}	316.0	316.8	165.9	167.5	39.3	43.7		
M _{IV}	274.4	274.2	139.0	137.5	23.8	29.2		0.4
M _V	260.7	261.2	133.2	132.9		28.8		
N _I	106.4	106.0	33.0	43.3	4.3	8.7		
N _{II}	92.5	93.5		36.0		5.9		
N _{III}	75.0	76.6		31.0				
N _{IV,V}	53.2	55.9*	17.2	18.1*		2.0*		
N _{VI, VII}	26.0	28.0*						

(Data are given in terms of ν_0/R where ν_0 is the frequency of the limit and R is the Rydberg constant.) The columns headed "Robinson" give Robinson's observations. Those headed "Spec." give the values of the levels determined from X-ray spectroscopy, as given by Siegbahn.

* Indicates mean value.

Robinson found on his plates a number of "lines" representing groups of photoelectrons. The corresponding absorption limits, calculated in the manner just described, are given in Table V; values determined spectroscopically are added for comparison. We note

that Robinson's results are *in complete agreement with the data of X-ray spectroscopy*. It is noteworthy that he found *one K level, three L levels, five M levels*, and, in the case of U 92, five of the seven *N levels*, the pairs of levels $N_{IV,V}$ and $N_{VI,VII}$ being too close together to be resolved in his apparatus.

The technique of measuring the velocities of photoelectrons ejected by X-rays has been developed to the point that this method is now one of the most reliable for the precise determination of fundamental constants.¹

197. The Scattering of X-rays.—In spite of the remarkable early success of the classical theory in accounting for many features of the scattering of X-rays, developments in this subject since 1920 illustrate strikingly the inadequacy of the classical theory and the superiority of quantum theory.

(a) *The Total Scattering Coefficient.*—When a beam of monochromatic X-rays of intensity I traverses a thickness dx of scattering material, part of the incident energy is scattered, for which we may write $\sigma I dx$. The coefficient σ thus defined is called the *total linear scattering coefficient* for X-rays of that wave length. According to the classical theory, on the two assumptions (1) that the electrons scatter as if they were free and (2) interference between the wavelets scattered by different electrons may be ignored, the value of σ can be found by multiplying the scattering coefficient σ_e for a single electron, as given in Eq. (263) in Sec. 184, by n , the number of electrons per cubic centimeter of the material. If we also divide by ρ , the density of the material, we find thus for the *classical mass scattering coefficient*

$$\frac{\sigma_0}{\rho} = \frac{n\sigma_e}{\rho} = \frac{8\pi}{3} \frac{e^4}{m^2 c^4} \frac{n}{\rho}. \quad (283)$$

Thus, if we assume that $\sigma = \sigma_0$, σ/ρ should be independent of wave length, and it should not vary greatly with the material, since n is roughly proportional to ρ (the number of electrons per atom being about half of the atomic weight); it should, in fact, be nearly equal to 0.2, as may be verified by putting into Eq. (283) the numerical values of the constants e , m , and c and the values of n and ρ for some scattering material, such as carbon or aluminum. In Fig. 169 are shown data on scattering for a number of elementary substances; the ordinates represent the ratio of the scattering coefficient σ for the substance, denoted by σ_s in the figure, to the value σ_0 as given in Eq. (283). We note from the figure that for light scattering materials and for X-rays

¹ See, e.g., KRETSCHMAR, *Phys. Rev.*, vol. 43, p. 417 (1933).

of $\lambda > 0.2$ angstrom the scattering coefficients approximate to the classical, free-electron value, σ_0 . Otherwise, there are wide departures.

For heavier materials, and especially at longer wave lengths, σ greatly exceeds σ_0 . This excessive scattering is easily explained classically. When the wave length becomes comparable with the distances between the electrons in the atom, they no longer scatter independently; the waves scattered by different electrons become

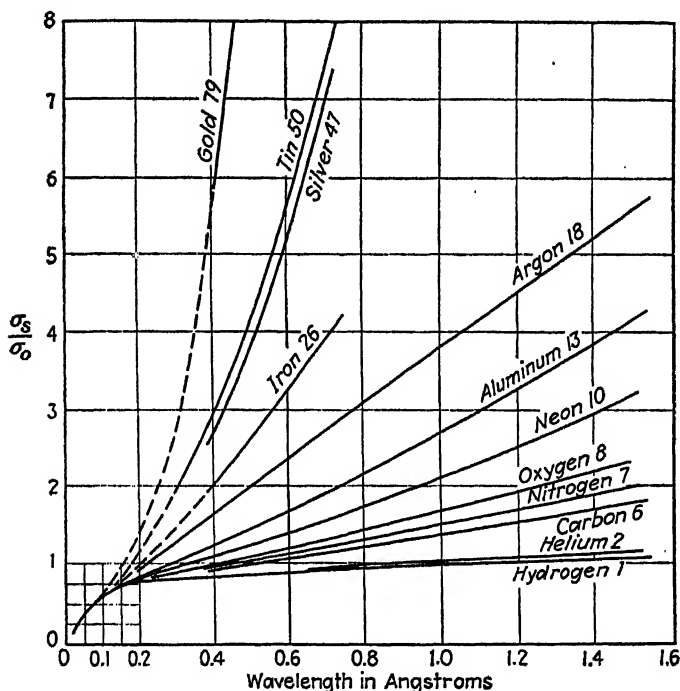


FIG. 169.—Scattering coefficients σ_s of various elements for different wave lengths, relative to the classical value σ_0 . (Reproduced by courtesy of D. Van Nostrand Company, Inc., from Fig. III-3 in A. H. Compton and S. K. Allison, "X-rays in Theory and Experiment," 1935.)

superposed on each other more or less in the same phase, and constructive interference occurs. If the wave length is actually long as compared to the distances between the N electrons in a given atom, the relative phase differences of the waves are all small, consequently the amplitude of the resultant scattered wave is approximately proportional to N , and the intensity of the scattered radiation is proportional to N^2 . For a heavy atom, the difference is enormous.

The decrease of σ at short wave lengths which is evident from Fig. 169, on the other hand, is impossible of explanation by classical

theory. For $\lambda = 0.1$ Å, the observed scattering coefficient is only about 0.15, instead of the value of about 0.2 deduced above. For γ -rays, the wave lengths of which are of the order of 0.02 Å or less, the total absorption coefficient of carbon, including photoelectric absorption as well as that due to scattering, is only about 0.06. For such short waves, however, the assumption that the electrons scatter independently, and approximately at the same rate as if they were free, should certainly be justified; for, on classical theory, they will not move far enough from their equilibrium positions to call appreci-

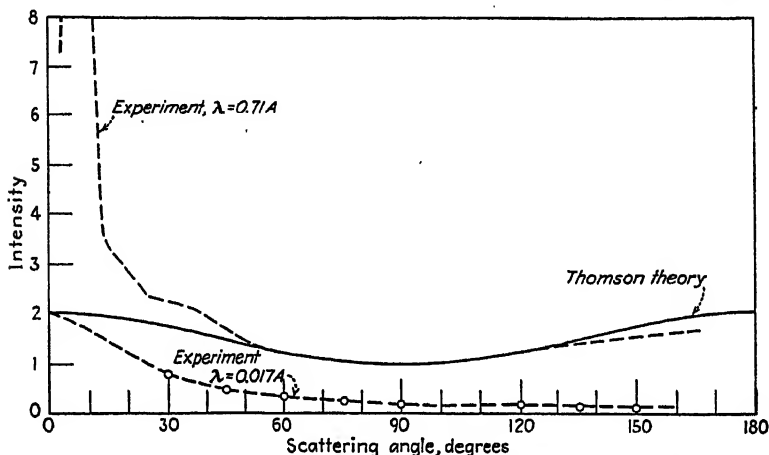


FIG. 170.—Relative scattered intensity at various angles of scattering, for X-rays of λ 0.71 Å scattered by mesitylene, and for X-rays of λ 0.017 Å scattered by iron. (Reproduced by courtesy of D. Van Nostrand Company, Inc., from Fig. III-4 in A. H. Compton and S. K. Allison, "X-rays in Theory and Experiment," 1935.)

able elastic forces into play. At this point classical theory fails badly. In the next section, a result will be quoted from wave mechanics which agrees very well with observations in this region of short wave lengths.

(b) *The Angular Distribution of Scattered X-ray Energy.*—The classical pulse theory of X-rays led to Eq. (265) in Sec. 184,

$$I_{\phi}' = \frac{I}{r^2} \frac{e^4}{m^2 c^4} \frac{1 + \cos^2 \phi}{2}$$

for the intensity I_{ϕ}' of X-rays scattered per electron at an angle ϕ from the direction of the incident beam, the intensity of which is I ; r is the distance between the scatterer and the place where the scattered energy is measured. According to this equation, the scattering should be symmetrical forward and backward about a plane passing through

the scatterer and perpendicular to the direction of the incident rays. If $\phi = 0$ or $\phi = \pi$, $1 + \cos^2\phi = 2$; if $\phi = \pi/2$, $1 + \cos^2\phi = 1$. Thus the energy scattered in the forward or backward directions should be twice that scattered at 90° from the incident beam.

In Fig. 170 are shown data on the variation with scattering angle ϕ of the intensity of X-rays of wave length $\lambda = 0.71$ Å scattered from liquid mesitylene $C_6H_3(CH_3)_3$. For comparison, the variation predicted by the classical theory of Thomson, as represented by the above formula, is shown by a solid line in the figure. We note that the intensity of scattering agrees well with the classical formula at scattering angles greater than 45° but shows a marked departure at smaller angles. Such a departure might perhaps be explained classically in the same way in which the excess scattering at long wave lengths was explained above. But in Fig. 170 data are also shown on the scattering of γ -rays of wave length 0.017 Å from iron. These data show much smaller scattering than is required by the classical formula (except, presumably, at a zero scattering angle), and the scattered radiation exhibits great dissymmetry, being confined almost entirely to forward directions.

Such dissymmetry in the angular distribution presents difficulties for any classical theory, but it is fully accounted for by the wave-mechanical theory.¹

198. The Compton Scattering of X-rays. (a) *The Compton Effect.* It was early found that, in general, secondary radiation is *less penetrating* than the primary—a fact which was later explained by the presence in the secondary beam of the *characteristic fluorescent* radiation, which, as we have seen above [Sec. 190(c)], is always of longer wave length and, therefore, is less penetrating than the exciting primary beam. In a scattering material of very low atomic number, such as carbon, however, the fluorescent radiation is of such a long wave length as to be absorbed by even a thin film of air. Yet, even with carbon as a scatterer, the secondary beam was found to be somewhat less penetrating and presumably, therefore, of somewhat longer mean wave length than the primary.

In a spectrometric investigation of scattered radiation, A. H. Compton² showed that when monochromatic radiation falls upon a scatterer, the scattered beam is composed of *two* lines, one corresponding in wave length to that of the primary beam and the other being of definitely *longer* wave length. Figure 171 shows the spectrum of scattered radiation from a carbon scatterer when irradiated by the $K\alpha$

¹ Cf. HETTLER, W., "The Quantum Theory of Radiation," p. 154, 1936.

² *Phys. Rev.*, vol. 21, p. 715; vol. 22, p. 409 (1923).

line of molybdenum, the scattered radiation being observed at an angle of 90° from the incident radiation. The vertical line aa gives the wave length of the primary radiation.

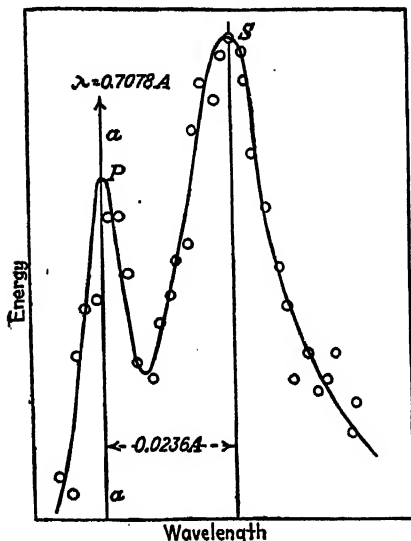


FIG. 171.—The spectrum of scattered X-rays, showing the unmodified line P and the modified or "shifted" line S (Compton).

There are seen to be two maxima, or lines, one of which, P , corresponds *exactly* to the primary radiation; the other, S , is "shifted" toward longer wave lengths from the position of the primary, the difference in wave length between the *shifted* and the *unshifted* components being 0.0236 angstrom.

In order to explain the occurrence of the shifted component, Compton boldly applied the extreme quantum picture of radiant energy, according to which a beam of radiation consists of a stream of localized quanta of energy, or photons. He assumed that the scattering process could be treated as an elastic collision between a photon and an electron and that this collision is governed by the

two laws of mechanics: (1) the conservation of energy and (2) the conservation of momentum.

Let an incident photon containing energy $E_0 = h\nu_0$ rebound from an electron of rest mass m , initially at rest at the point a (Fig. 172).

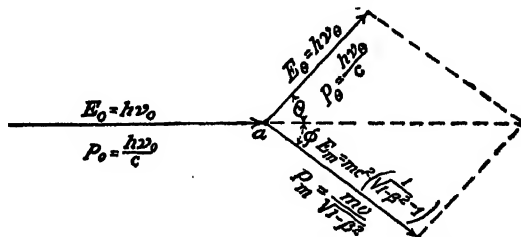


FIG. 172.—Vector diagram showing the conservation of momentum when a photon is scattered by an electron.

After the collision, the electron will move in a direction making an angle of ϕ , say, with the initial direction of motion of the photon while the photon itself will move in a direction making an angle θ with its initial direction. Qualitatively, one sees, at once, that the energy

E_θ of the photon after collision and, therefore, also the frequency must be less than that of the photon before collision, since some of the energy of the incident photon must have been given to the electron if it has been set in motion.

If E_m is the kinetic energy given to the electron as the result of the collision, we have, from the law of the conservation of energy,

$$E_0 = E_\theta + E_m.$$

Using the relativistic formula, Eq. (68) in Sec. 68, we have for the kinetic energy of the electron

$$E_m = mc^2 \left(\frac{1}{\sqrt{1 - \beta^2}} - 1 \right), \quad \beta = \frac{v}{c} \quad (284)$$

where m is its rest mass, v is its velocity after the collision, and c is the speed of light. Inserting this value of E_m in the preceding equation, and also $E_0 = h\nu_0$, $E_\theta = h\nu_\theta$, we have

$$h\nu_0 = h\nu_\theta + mc^2 \left(\frac{1}{\sqrt{1 - \beta^2}} - 1 \right). \quad (285)$$

Furthermore, as stated in Sec. 35, a beam of radiation in free space carries momentum equal to the energy carried by it divided by c . Hence a photon, containing energy $h\nu$, carries momentum $h\nu/c$. The momentum of an electron moving with speed v is given by Eq. (67) in Sec. 67. Thus for the momenta we have:

$$\begin{aligned} h\nu_0/c &= \text{the momentum of the photon before the collision,} \\ h\nu_\theta/c &= \text{the momentum of the photon after the collision,} \\ mv/\sqrt{1 - \beta^2} &= \text{the momentum of the electron (after the collision).} \end{aligned}$$

From the law of the conservation of momentum, therefore, we can write for the x -component of momentum (in the direction of propagation of the incident photon)

$$\frac{h\nu_0}{c} = \frac{h\nu_\theta}{c} \cos \theta + \frac{m\beta c}{\sqrt{1 - \beta^2}} \cos \phi \quad (286)$$

and for the y -component, in the plane containing the directions of motion of both the scattered photon and the electron,

$$0 = \frac{h\nu_\theta}{c} \sin \theta - \frac{m\beta c}{\sqrt{1 - \beta^2}} \sin \phi. \quad (287)$$

In the three independent equations (285), (286), and (287), we have the four unknown quantities ν_θ , θ , ϕ , and β , assuming ν_0 , the frequency of the incident photon, to be known. By simultaneous solution, we can

obtain three of them in terms of a fourth, say θ , which is the direction with respect to the incident beam in which the photon is "scattered" as a result of the particular collision.

If we introduce the wave lengths of the 2 photons (whatever wave length may mean in terms of this picture!), $\lambda_0 = c/\nu_0$ and $\lambda_\theta = c/\nu_\theta$, we can write the last two equations thus:

$$\frac{h}{\lambda_0} - \frac{h}{\lambda_\theta} \cos \theta = \frac{m\beta c}{\sqrt{1-\beta^2}} \cos \varphi, \quad \frac{h}{\lambda_\theta} \sin \theta = \frac{m\beta c}{\sqrt{1-\beta^2}} \sin \varphi.$$

Squaring these two equations and adding them, we have

$$\frac{h^2}{\lambda_0^2} + \frac{h^2}{\lambda_\theta^2} - \frac{2h^2}{\lambda_0\lambda_\theta} \cos \theta = \frac{m^2\beta^2 c^2}{1-\beta^2} = \frac{m^2 c^2}{1-\beta^2} - m^2 c^2. \quad (288)$$

Similarly Eq. (285) can be written, after dividing it through by c ,

$$\frac{h}{\lambda_0} - \frac{h}{\lambda_\theta} + mc = \frac{mc}{\sqrt{1-\beta^2}}.$$

$$\therefore \frac{h^2}{\lambda_0^2} + \frac{h^2}{\lambda_\theta^2} - \frac{2h^2}{\lambda_0\lambda_\theta} + 2mch \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_\theta} \right) + m^2 c^2 = \frac{m^2 c^2}{1-\beta^2}.$$

Subtracting Eq. (288) from the last equation, we obtain

$$\frac{2h^2}{\lambda_0\lambda_\theta} (\cos \theta - 1) + 2mch \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_\theta} \right) = 0.$$

$$\therefore \lambda_\theta - \lambda_0 = \frac{h}{mc} (1 - \cos \theta). \quad (289a)$$

If we insert here the values $h = 6.61 \times 10^{-27}$, $m = 0.9107 \times 10^{-27}$, $c = 2.998 \times 10^{10}$, and multiply by 10^8 , we have

$$\lambda_\theta - \lambda_0 = 0.0242 (1 - \cos \theta) \text{ angstroms.} \quad (289b)$$

This important equation states that *when incident radiation of wave length λ_0 is scattered by a free electron at an angle θ , the wave length λ_θ of the scattered radiation should be greater than that of the incident radiation by the quantity $0.0242(1 - \cos \theta)$, which, for a given angle θ , is constant whatever the incident wave length.* When the scattering angle is 90° , the "shift" in wave length between the primary and the scattered beam should be 0.0242 angstrom. This is very nearly in accord with Compton's measurements, quoted above. Further, the shift should be entirely independent of the material of the scatterer, and for various angles θ of scattering it should be proportional to $(1 - \cos \theta)$. These predictions have been abundantly confirmed by

the investigations of Compton and his collaborators,¹ Ross,² J. A. Becker,³ Allison and Duane,⁴ and others. For details, the student is referred to the original articles.

From Eqs. (284), (285), and (289a) it is readily found that the energy E_m of the recoiling electron is

$$E_m = h\nu_0 \frac{\zeta(1 - \cos \theta)}{1 + \zeta(1 - \cos \theta)}, \quad \zeta = \frac{h\nu_0}{mc^2};$$

and, from the construction of Fig. 172, it is evident that the electron must move in the forward direction, *i.e.*, $\phi < \pi/2$. Following Compton's prediction of their existence, these "recoil" electrons, as they are called, were discovered by C. T. R. Wilson⁵ and W. Bothe.⁶ They have been studied quantitatively by Compton and Simon⁷ and others. Bless,⁸ by the method of the magnetic "spectrograph," similar to Robinson's method for photoelectrons (Sec. 196), has measured the value of E_m for the recoil electrons and has found that the results are in agreement with the last equation.

(b) *The Unmodified Line.*—The presence of the so-called unmodified line (P in Fig. 171) may be accounted for as follows. In setting up our equations, it was assumed that the electron with which the photon collides is free. This assumption is justified if the energy given to the electron is much larger than the work necessary to detach it from the atom. It is possible, however, that another type of collision may occur in which the electron remains bound to the atom. Such a collision may be regarded as taking place between the photon and the atom as a whole, the mass of which is far greater than that of the electron. If the mass of the atom be substituted for m in Eq. (289a), it is readily seen that the computed change of wave length is so small as to be beyond the possibility of detection. It may well be that, in the scattering of photons, the electron sometimes behaves as if bound to the atom and sometimes as if free,—hence the unmodified and the modified scattered lines.

¹ *Phys. Rev.*, vol. 21, pp. 207, 483, 715 (1923); *Nat. Acad. Sci., Proc.*, vol. 10, p. 271 (1924).

² *Nat. Acad. Sci., Proc.*, vol. 9, p. 246 (1923); *Phys. Rev.*, vol. 22, p. 524 (1923).

³ *Nat. Acad. Sci., Proc.*, vol. 10, p. 342 (1924).

⁴ *Nat. Acad. Sci., Proc.*, vol. 11, p. 25 (1925); Woo, Y. H., *Phys. Rev.*, vol. 27, p. 242 (1926).

⁵ *Roy. Soc., Proc.*, vol. 104, p. 1 (1923).

⁶ *Zeits. f. Physik*, vol. 20, p. 237 (1923).

⁷ *Phys. Rev.*, vol. 25, p. 306 (1925).

⁸ Dissertation, Cornell University (1927); *Phys. Rev.*, vol. 29, p. 918 (1927).

It must be admitted, however, that the simple picture which we have employed above does not harmonize very well with the occurrence of an *unmodified* line. Nor is it clear how the *modified* line is affected by the presence of the nuclear field that normally holds the electron within the atom. These matters, like the properties of atomic states, can only be cleared up (?) by an application of wave mechanics.

The treatment of the impact of X-rays on atoms by wave mechanics leads to the following conclusions,¹ which, as far as tested, are confirmed by experiment. At low frequencies, the scattered radiation has the same frequency as the frequency ν of the incident radiation, constituting an unmodified scattered line. As the wave length becomes comparable with the dimensions of the electronic shells in the atoms, however, the intensity of the radiation thus scattered diminishes, especially at the larger angles of scattering. Furthermore, as ν becomes larger than the absorption limits for any atomic shell, Compton scattering accompanied by the ejection of recoil electrons out of this shell sets in; and, at considerably greater frequencies, the modified line due to this type of scattering becomes easily observable.

The Compton or modified line is very broad at comparatively low frequencies. This breadth can be thought of as caused by the motion of the electrons in the atom; wave mechanics gives a certain probability distribution for their velocities, from which the broadening can be computed. In our simple deduction of the Compton effect, the electron was assumed to be initially at rest; if it is assumed to have a component of velocity, positive or negative, in the direction of the incident radiation, the wave-length shift is different.²

As ν is then increased further, the Compton line becomes narrower. Eventually, in any direction of scattering other than that of the incident beam, the unmodified line becomes weaker than the Compton line, sooner at large angles of scattering than at small angles, and sooner for heavy atoms than for light ones. Finally, only the modified line remains in appreciable intensity; this happens, for example, in light scattered at right angles to the incident beam, for $\nu/\nu_K > 300$ in the case of carbon or $\nu/\nu_K > 50$ in the case of lead, ν_K being the absorption limit for the *K* shell. The modified line is then fairly sharp and approximately the same in all respects as if the electrons in the atom were free (and at rest). According to the theory, the modified line should exhibit the same features as to polarization as the unmodified line, except at extremely high frequencies.

¹ Cf. BLOCH, *Phys. Rev.*, vol. 46, p. 674 (1934); HEITLER, *op. cit.*

² For observations see DuMOND and KIRKPATRICK, *Phys. Rev.*, vol. 52, p. 419 (1937).

According to the theory, furthermore, the other process by which X-rays may eject electrons from atoms, that of photoelectric absorption, should likewise decrease rapidly with increasing frequency of the incident radiation and should be relatively less important in light atoms than in heavy ones; the probability of the ejection of a photoelectron should become equal to that of the ejection of a Compton electron at about $\lambda = 0.5 \text{ \AA}$ or $\nu/\nu_K = 85$ in carbon, or $\lambda = 0.025 \text{ \AA}$ or $\nu/\nu_K = 6$ in lead. Thus at high frequencies the ejection of Compton recoil electrons should constitute effectively the only form of action of radiation upon the electrons in matter.

It may be remarked that only the *unmodified* scattered radiation contributes to the beams that are reflected (or more accurately, diffracted) in certain characteristic directions by crystals (Sec. 186) or gratings (Sec. 200) and upon which X-ray spectroscopy depends. Only to this part of the radiation can we apply the classical conception of scattered waves which are coherent with the incident waves, so that the waves scattered by different atoms possess definite relations of phase to one other and are capable of interfering constructively. The waves representing the Compton scattering must be supposed to be completely incoherent.

(c) *The Scattering Coefficient as a Function of Frequency.*—The wave-mechanical theory gives a good account of the observed decrease of the scattering coefficient at high frequencies, which was described in the last section. At frequencies much above the K absorption limit of the atom, as has been said, we can treat the electrons in the atom as if they were free electrons, scattering independently by the Compton process. Wave mechanics leads to the following formula for the amount of energy scattered per electron per second by a group of free electrons:¹

$$\Omega = \frac{8\pi}{3} \frac{c^4}{m^2 c^4} I \cdot \frac{3}{4} \left\{ \frac{1 + \zeta}{\zeta^2} \left[\frac{2(1 + \zeta)}{1 + 2\zeta} - \frac{1}{\zeta} \log(1 + 2\zeta) \right] + \frac{1}{2\zeta} \log(1 + 2\zeta) - \frac{1 + 3\zeta}{(1 + 2\zeta)^2} \right\}, \quad (290a)$$

where I is the intensity of the incident beam of frequency ν and

$$\zeta = \frac{h\nu}{mc^2}.$$

(h = Planck's constant, m = rest mass of electron, c = speed of light). As $\zeta \rightarrow 0$, the complicated quantity in braces approaches $\frac{4}{3}$ and Ω

¹ KLEIN and NISHINA, *Zeits. f. Physik*, vol. 52, p. 853 (1929).

approaches the classical value as found by Thomson [cf. Eq. (262) in Sec. 184(a)]. If ξ becomes large, however, as it is for hard γ -rays, approximately

$$\Omega = \frac{\pi e^4 I}{h \nu m c^2} \left(\log \frac{2 h \nu}{m c^2} + \frac{1}{2} \right), \quad (290b)$$

which decreases toward zero with increasing ν , varying nearly as $1/\nu$. In any case, the scattering coefficient is

$$\sigma = \frac{n \Omega}{I},$$

where n is the number of electrons per unit volume. This formula fits the existing data very well.

(d) *The Compton Scattering Process.*—It may be of interest to remark that even the *classical* theory leads us to expect radiation that is scattered by *free* electrons to be of somewhat *longer wave length* than the incident beam. For the electrons must take up the momentum that is carried by that part of the incident radiation that is scattered, since the scattered waves, because of the symmetry of their distribution, carry away no momentum on the whole. The electrons will thus be set in motion in the direction of propagation of the incident beam (an effect which can be ascribed to radiation pressure), and the radiation scattered by them will be decreased in wave length because of the Doppler effect. It can even be shown that, if one quantum $h\nu$ is scattered by a free electron initially at rest, the average increase in wave length is of the same order as the Compton shift. But classical theory cannot explain the occurrence of a *sharp line* in the Compton effect. Furthermore, the moving electron would have to move an impossible distance in the scattering substance. Quantum theory is superior in that the entire photon is scattered as a unit by a single electron, and only this one electron is set into motion.

Finally, reference should be made to another experiment, performed much later, which further heightened the contrast between classical and quantum theory by demonstrating that the recoil electron and the scattered X-ray photon are produced *at the same time*.¹ The scattered photon was detected by observing in a cloud chamber the track of a photoelectron released by it at some other point in the scattering material, which was a gas. Recoil tracks and photoelectric tracks due to scattered X-rays were observed to occur simultaneously. This observation shows very clearly the discontinuous nature of the scattering process, which, according to classical theory, should be

¹ BOTHE and GEIGER, *Zeits. f. Physik*, vol. 26, p. 44 (1924); vol. 32, p. 639 (1925).

continuous. Most of the time nothing is happening in the scattering material. At irregular instants, however, a recoil electron is projected from one of the molecules, and almost simultaneously (presumably later by the amount of time it takes light to pass from one point to the other) a photoelectron is ejected from another molecule. At such instants, we say that an X-ray photon has been "scattered" by the first molecule and absorbed by the second one.

199. The Refraction of X-rays. (a) *Refraction and Bragg's Law.*—The discovery of the refraction of X-rays and the subsequent measurement of indices of refraction are excellent illustrations of the great advances which have been made in the technique of X-ray measurements since the discovery of the action of the crystal grating. The

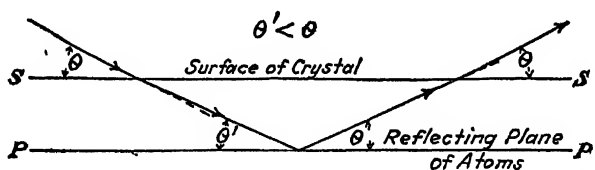


FIG. 173.—Refraction of a beam of X-rays entering the surface of a crystal.

first positive evidence that X-rays are measurably refracted came from the work of Stenström,¹ who showed from accurate measurements of wave length that Bragg's law of the reflection of X-rays from crystals does not yield identical values when the wave length of a given line is computed from different orders of reflection. Hjalmar² found, for example, that the apparent wave length of the Fe $K\alpha_1$ line as measured in the first order by reflection from a gypsum crystal ($2d = 15.155$ angstroms) was 1.9341 Å; while measurements in the sixth order gave 1.9306 Å, nearly 0.2 percent less.

This apparent failure of the Bragg formula was shown to be due to the refraction of the beam of X-rays as it entered the crystal, the deviation of the beam being such as to indicate that the index of refraction is less than unity. In Fig. 173 is shown the path of a ray as it enters the surface SS of a crystal at a glancing angle θ and is incident on the Bragg plane PP at an angle θ' . Bragg's law in the form

$$n\lambda' = 2d \sin \theta',$$

where λ' is the wave length *in the crystal*, gives the true law of reflection *at the crystal plane*. Measurements give θ , however, and it is λ , the wave length in *air*, that is desired. By the ordinary laws of optics

¹ Dissertation, Lund (1919).

² *Zeits. f. Physik*, vol. 15, p. 65 (1923).

$$\mu = \frac{\lambda}{\lambda'} = \frac{\cos \theta}{\cos \theta'}$$

Substituting values of λ' and of $\sin \theta' = (1 - \cos^2 \theta')^{1/2}$ in the preceding equation, we find $n\lambda/\mu = 2d (1 - \cos^2 \theta/\mu^2)^{1/2}$, whence

$$\begin{aligned} n\lambda &= 2d (\mu^2 - \cos^2 \theta)^{1/2} = 2d (\mu^2 - 1 + \sin^2 \theta)^{1/2} \\ &= 2d \sin \theta \left[1 + \frac{(\mu - 1)(\mu + 1)}{\sin^2 \theta} \right]^{1/2}. \end{aligned}$$

Since $1 - \mu$ is extremely small for X-rays, we can replace $\mu + 1$ by 2 in this expression and then expand in powers of $1 - \mu$ and keep only the first power. Thus we obtain

$$n\lambda = 2d \sin \theta \left(1 - \frac{1 - \mu}{\sin^2 \theta} \right). \quad (291)$$

A few of the values of $\delta = 1 - \mu$ found in this way for mica and calcite by Larsson¹ are as follows:

λ , Å.....	1.537	2.499	3.447	5.166	7.111	8.320
$\delta \times 10^6$, mica.....	8.94	24.6	49.1	103	182	262
$\delta \times 10^6$, calcite.....	8.8	22.4	41.9			

(b) *Theory of the Refraction of X-rays.*—Classical theory gives for the index of refraction μ of a slightly refracting and absorbing substance, for waves of frequency ν ,

$$\mu^2 = 1 + \frac{e^2}{\pi m} \sum_i n_i \frac{\nu_i^2 - \nu^2}{(\nu_i - \nu)^2 + \gamma_i^2 \nu^2},$$

where n_i is the number of electrons per unit volume which have a natural frequency ν_i (of undamped vibration), γ_i is a constant representing the effect of damping, and the summation is to be extended to include all of the different groups of electrons in the refracting material.² If ν is very much greater than all of the characteristic frequencies ν_i , and all γ_i 's are small, we may neglect ν_i^2 and $\gamma_i^2 \nu^2$ in comparison with ν^2 ; if also μ differs very little from 1, we may write $\mu^2 = [1 + (\mu - 1)]^2 = 1 + 2(\mu - 1)$, approximately. Thus for very large ν the last formula can be written:

$$\mu = 1 - \delta, \quad \delta = \frac{ne^2}{2\pi m \nu^2}, \quad (292a)$$

where $n = \sum_i n_i$ and represents the total number of electrons per unit

¹ Quoted in SIEGBAHN, *loc. cit.*

² See WOOD, R. W., "Physical Optics," 3d ed., p. 488, 1936.

volume. Or, substituting $\nu = c/\lambda$, we have

$$\frac{\delta}{\lambda^2} = \frac{ne^2}{2\pi mc^2}. \quad (292b)$$

Thus, for frequencies much greater than any characteristic frequencies of the refracting substance, the quantity δ/λ^2 should be independent of the incident frequency.

Analogous results are obtained from wave mechanics. In place of n_i , a coefficient occurs that is mostly less than 1. The frequencies ν_i correspond to the various transitions that the atom can undergo either into a higher quantum state of the neutral atom or into an ionized state as the result of a process of photoelectric absorption. The terms corresponding to photoelectric absorption, however, have an

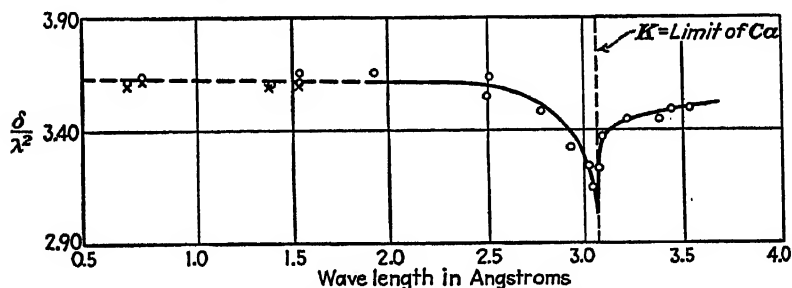


FIG. 174.—Values of δ/λ^2 for calcite, showing anomalous dispersion near the Ca K absorption limit. [\times = data by Bearden and Shaw; \circ = data by Larsson, Thesis, Univ. Uppsala, Ann. Reports, (1929).]

appreciable effect upon the dispersion only when the incident frequency ν is close to one of the absorption limits of the atom. At other frequencies, Eq. (292b) should hold.

Experiment reveals a behavior of the refractive index in good agreement with these theoretical predictions. For example, in Fig. 174 is shown the observed variation of δ/λ^2 with λ for calcite in the neighborhood of the K absorption limit of Ca at 3.064 Å. The dotted portion of the curve represents the value $\delta/\lambda^2 = 3.67$, computed for calcite from Eq. (292b). The equation is seen to hold very well from 0.5 to 2.5 Å.

(c) *Direct Measurement with a Prism.*—Although the index of refraction for X-rays differs only slightly from unity, its direct measurement by a prism has been accomplished by several observers. A right-angled prism may be used, the X-rays entering one face at a glancing angle of only a few minutes of arc and emerging almost perpendicularly from the opposite face; the prism may be substituted for the mirror *M* in the arrangement shown in Fig. 175, which is described in the next

section. With modern technique, the accuracy of such methods may be very high; the prism provides, in fact, a method for the measurement of absolute X-ray wave lengths with a precision of 1 part in 10,000.* The wave lengths are calculated by means of the complete formula furnished by wave mechanics. The fact that wave lengths so measured agree with the same wave lengths as determined by other methods tends to confirm the correctness of the wave-mechanical theory of X-ray dispersion.

(d) *Measurement by Total Reflection.*—Since the index of refraction of a material for X-rays is less than unity, a beam of X-rays incident on a polished surface at a sufficiently small glancing angle should be totally reflected. According to the usual law of optics, this should occur for *glancing* angles θ_R , or less, where

$$\cos \theta_R = \mu = 1 - \delta, \quad \sin \theta_R = \theta_R = \sqrt{2\delta}$$

for small values of θ_R . Since, for a given substance, δ/λ^2 is approximately constant, it follows that, to the same approximation, $\theta_R \propto \lambda$. From the value of δ determined by methods described above, θ_R may be computed. Several representative values of θ_R are shown in Table VI.

TABLE VI.—SOME VALUES OF THE CRITICAL ANGLE OF TOTAL REFLECTION θ_R

Substance	λ , angstroms	δ	θ_R
Glass.....	0.7078	1.64×10^{-8}	6' 10''
Calcite.....	1.537	8.80	14' 25''
Calcite.....	3.734	49.2	34' 5''
Quartz.....	10.0	356	91' 40''

Reversing the procedure, one may observe θ_R and compute δ .

One of the first measurements of X-ray refraction was by A. H. Compton,¹ using this method. The measurements of Doan² are instructive. Doan's apparatus, following Compton's method, is shown schematically in Fig. 175. A beam of X-rays from a target T , after passing through a slit S_1 , falls onto a crystal K , by means of which one of the characteristic lines in the spectrum of T may be reflected at a very small glancing angle onto the mirror M . With a sufficiently small angle of incidence, this beam is totally reflected from M and falls on a photographic plate PP at R . The mirror M

* BEARDEN and SHAW, *Phys. Rev.*, vol. 46, p. 759 (1934); BEARDEN, *Phys. Rev.*, vol. 54, p. 698 (1938).

¹ *National Research Council, Bull.* 20, p. 50 (1922).

² *Phil. Mag.*, vol. 4, p. 100 (1927).

is turned very slowly during an exposure; and, when the critical angle is reached, reflection ceases. This critical angle can be determined by noting the distance on the plate between the point *O*, which indicates the position of the direct beam, and the extreme edge of the record of the reflected beam. The metals under study were sputtered onto a highly polished optical glass surface.

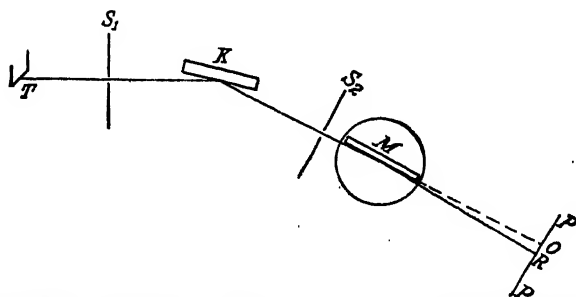


FIG. 175.—Doan's apparatus for measuring the index of refraction of X-rays by the method of total reflection.

Table VII shows some of the results obtained by Doan.

TABLE VII.—INDEX OF REFRACTION OF X-RAYS BY TOTAL REFLECTION

Wave length, angstroms	Substance	Critical angle	$\delta = 1 - \mu, \times 10^6$
0.7078	Ni	10' 15"	4.42
0.7078	Ag	11' 42"	5.76
1.389	Ni	16' 9"	10.98
1.389	Cu	19' 36"	16.2
1.537	Ni	24' 40"	25.5
1.537	Cu	20' 24"	17.7
1.537	Ag	26' 42"	30
1.537	Au	31' 24"	41.6

The accuracy of this method is not high, however, since the critical angle of total reflection for any given wave length is sharply defined only if the coefficient of absorption for that wave length is negligible—which is by no means the case with X-rays, particularly of the longer wave lengths. For example, the value of δ for Cu $K\alpha_1$ radiation ($\lambda = 1.537$ Å) reflected by silver is approximately 30×10^{-6} , from which $\theta_E = \sqrt{2\delta} = 27$ minutes of arc (about). Were there no absorption, the coefficient of reflection for glancing angles up to 27 minutes of arc would be as shown by the dotted line in Fig. 176. Nähring,¹

¹ *Phys. Zeits.*, vol. 31, p. 799 (1930).

however, shows that, using classical theory and the known coefficient of absorption of silver for $\lambda = 1.537$ Å, the coefficient of reflection R for various glancing angles of incidence θ should be as given by the full line of Fig. 176. The dots represent Nähring's observations. We note that, although $R > 0.5$ for $\theta < 25'$, $R = 1$ only at $\theta = 0$.

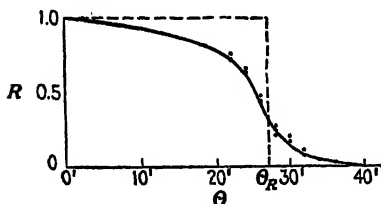


FIG. 176.—Coefficient of reflection R of silver for $\text{Cu } K\alpha_1$ radiation ($\lambda = 1.537$ angstroms) at various glancing angles of incidence θ .

200. Measurement of X-ray Wave Lengths by a Ruled Grating.

With the discovery of the total reflection of X-rays, it became plausible that a ruled grating might be used to measure X-ray wave lengths in exactly the same way that a grating is similarly used in the optical region, provided that the glancing angle between the X-ray beam and the ruled surface is less

than the critical angle for total reflection. Compton and Doan¹ were the first to make measurements of this kind. Using a grating of speculum metal with 50 lines per millimeter,² they found the wave length of the $K\alpha_1$ line of molybdenum to be $\lambda = 0.707 \pm 0.003$ angstrom.

For a description of the development of the technique of using ruled gratings for the measurement of X-ray wave lengths, the reader is referred to original articles.³ We have already mentioned (Sec. 187) the precise work of Bearden,⁴ who measured the wave length of the $\text{Cu } K\alpha_1$ line (1.5406 Å) with gratings having 100 or 300 lines per mm. ruled on glass sputtered with gold. His results are probably correct within less than 0.01 per cent.

For several years, there existed a discrepancy of 0.2 to 0.3 percent between ruled-grating and crystal values of X-ray wave lengths. As stated in Sec. 187, however, this discrepancy appears now to have been removed by the discovery of an error in the crystal values, due in turn to an error in the value assumed for the electronic charge. Measurements of X-ray wave lengths made with ruled gratings seem to be the most reliable ones, since they involve no assumptions as to the homogeneity of a crystal; in fact, the only elements that enter into the

¹ *Nat. Acad. Sci., Proc.*, vol. 11, p. 598 (1925).

² The arrangement used by Compton and Doan may be indicated schematically by replacing the mirror M , Fig. 175, by the grating. The glancing angle was less than 25 minutes of arc.

³ THIBAUD, *J. de Physique et le Radium*, vol. 8, pp. 13, 447 (1927); BÄCKLÉN, Dissertation, Uppsala (1928); BEARDEN, *Phys. Rev.*, vol. 33, p. 1088 (1929); HOWE, *Rev. Sci. Instruments*, vol. 1, p. 749 (1930).

⁴ *Phys. Rev.*, vol. 48, p. 385 (1935).

determination by this method are the wave theory of light as propagated in a vacuum and such well-tested operations as the measurement of angles and the counting of lines under a micrometer microscope.

SOME RECENT DEVELOPMENTS IN X-RAY SPECTROSCOPY

Within the scope of this book, we cannot give the "latest word" on any of the various ramifications of the subjects introduced. The selection of a few topics for very cursory discussion in this last part of the chapter on X-rays is, perforce, arbitrary. For an acquaintance with the many other interesting lines of current research in this field, the student is referred to the literature.

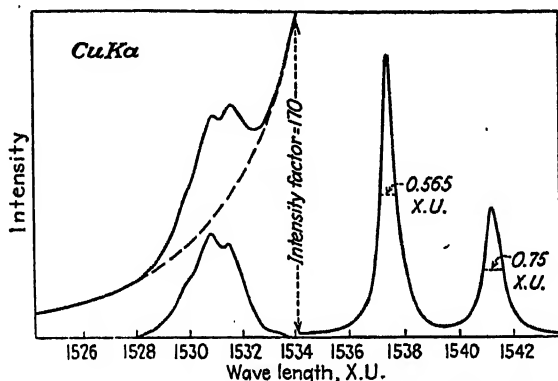


FIG. 177.—The $K\alpha$ doublet of Cu (29) with accompanying satellites. The left-hand half of the curve is plotted on a scale 170 times larger than the right-hand half, and below it is plotted the intensity due to the satellites alone, obtained by subtracting the presumable intensity due to other lines as represented by the dotted curve.

201. Multiple Ionization of Inner Electron Shells.—Our knowledge of the inner structure of the atom has been advanced greatly by work on such X-ray phenomena as satellites, line widths, and relative intensities of lines and the theoretical explanations of these phenomena. The explanations all involve atomic states characterized by the absence of 2 or more inner electrons.¹

(a) *Satellites or Second-order Lines.*—It was perhaps fortunate that the spectral apparatus available to Moseley and the early workers did not have the sensitivity and resolving power of present-day spectrometers. The lines which they observed were the more intense and more easily resolved lines of X-ray spectra. These are the lines which are represented on the conventional energy-level diagram (Sec. 191) and were readily interpreted in terms of the extended

¹ For a more detailed discussion than that given here see RICHTMYER, *Rev. Modern Physics*, vol. 9, p. 391 (1937).

Bohr theory of atomic spectra. As we have seen, these lines are due to transitions between states of *single ionization*. They may be called *first-order* lines.

With improvements in technique, many other lines were subsequently discovered which did not fit into the conventional diagram. The majority of these lines were rather faint, were usually found close to and on the short-wave-length side of the more intense "diagram" lines, and hence were called "satellite" lines. A typical spectral curve of the satellite structure accompanying the $K\alpha$ lines of copper is reproduced¹ in Fig. 177. The satellites accompanying the $L\beta_1$, $L\beta_2$, and $L\gamma_1$ lines for Ag (47) are designated by primes in Fig. 158. As illustrated in these figures, the satellite structure is observed to be very complex, containing numerous component lines of various intensities. Most (if not all) first-order lines are accompanied by such satellite structure, varying from line to line as well as from element to element. The total number of component satellite lines now known far exceeds the total number of diagram lines. If such a confusing array of lines had been presented to Moseley and the other early workers, it is probable that progress in X-ray spectroscopy would have been much less rapid.

In seeking a plausible interpretation of the origin of satellite lines, we may focus our attention upon such key characteristics as the following: (1) the minimum excitation voltage, or, if the lines are observed in fluorescence, the minimum photon energy for excitation; (2) wave-length positions of the lines; (3) variation in the relative intensities of the satellite lines with the atomic number of the radiating material, or with the voltage and current in the X-ray tube.

Because of the low intensity of the satellites, reliable experimental information on such characteristics is extremely difficult to obtain. After several trials, however, the excitation voltage of certain satellites was definitely shown to be somewhat greater than the excitation voltage of the accompanying first-order or parent line.² In the case of the type of K satellites shown in Fig. 177, the energy of excitation is found to be equal to the energy required to eject a K electron and *in addition* an L electron from the atom, the energies required to eject these electrons being calculated from absorption limits. Hence, we may assume, as a working hypothesis, that the initial state for the emission of these satellite lines is a state of *double ionization*, in which

¹ PARRATT, *Phys. Rev.*, vol. 50, p. 1, (1936).

² See for example, DRUYVESTEYN, *Zeits. f. Physik*, vol. 43, p. 707 (1927); PARRATT, *Phys. Rev.*, vol. 49, p. 132 (1934); COSTER, KUIPERS, and HUIZINGA, *Physica*, vol. 2, p. 870 (1935).

the atom has an electronic vacancy in both the K shell and the \bar{L} shell. Such a state may be called a " KL atomic state." In a similar way, other states of double ionization, such as KK , KM , LM , etc., should be possible.¹

An atom in a KL state may undergo a radiative transition into any one of a number of other states of double ionization, e.g., $KL \rightarrow KM$ (an electron dropping from the M shell into the L shell), or $KL \rightarrow LL$ (an electron dropping from the L shell into the K shell). Estimates of the atomic energy indicate that the loss of energy in the transition $KL \rightarrow LL$ should be slightly greater than in the diagram transition $K \rightarrow L$, which gives rise to the $K\alpha$ lines; hence, the former transition should give rise to satellites close on the short-wave side of the $K\alpha$ lines. Similarly, the transition $KL \rightarrow LM$ should give rise to satellites on the short-wave side of the $K\beta$ lines. In a doubly ionized atom, the two vacancies would function spectroscopically in much the same way as would 2 valence electrons (Secs. 150 to 152). Furthermore, the vacancies may have different l values corresponding to the various subshells; for example, the satellites associated with the $L\alpha$ diagram line are assumed to originate in transitions between the levels L_{III} , M_{IV} or L_{III} , M_V (written compactly as $L_{III}M_{IV,V}$) and the levels $M_{IV}M_{IV}$, $M_{IV}M_V$, M_VM_V (or $M_{IV,V}M_{IV,V}$).

Thus, we may explain the multiplicity of the observed satellite structures. As far as it has been possible to make approximate theoretical calculations, the predictions of wave mechanics lend support to the explanation just described for the origin of satellites.²

Some satellites have been reported which are believed to originate in transitions between states of *triply ionized* atoms. Thus, we may have satellite lines of second and higher orders, in analogy with the various higher orders of optical "spark" spectra. The construction of an energy-level diagram for second (or higher) order lines in X-ray spectra is extremely difficult, however, and, in general, has not been accomplished either empirically or theoretically. It should be stated, furthermore, that the theory of satellites has scarcely advanced beyond the qualitative stage, and it is entirely possible that some of these lines originate in some type of atomic process other than that just described.³

¹ This theory of satellites was proposed by Wentzel, *Ann. d. Physik*, vol. 66, p. 437 (1921) and modified by Druyvesteyn, *Zeits. f. Physik*, vol. 43, p. 707 (1927).

² For discussion and references, see Richtmyer and Ramberg, *Phys. Rev.*, vol. 51, p. 925 (1937).

³ See, e.g., RICHMYER, *Frank. Inst. J.*, vol. 208, p. 325 (1929); BLOCH, *Phys. Rev.*, vol. 48, p. 187 (1935).

(b) *The Auger Effect.*—In the discussion of satellites, we assumed the atom to be doubly ionized without considering how this state might be brought about. Presumably, it is possible for a cathode-ray electron to eject 2 electrons at once from an atom. If this is the origin of the doubly ionized atoms, theoretical estimates indicate that the intensity of satellites relative to the parent lines should decrease in a continuous manner with increasing atomic number. This relation is found by experiment to be true for *K* satellites, *i.e.*, those accompanying lines of the *K* series, but not for *L* or *M* satellites. *The intensity of the satellites accompanying the La line, for example, is observed to decrease rather abruptly as the atomic number increases from 47 to 50 and to increase rather abruptly at about 75; between atomic numbers 50 and 75, La satellites are practically unobservable.* This anomalous behavior as to intensity prevented for a time the universal acceptance of the Wentzel-Dryvesteyn theory of satellite origin. The difficulty was resolved when Coster and Kronig pointed out the importance, in this connection, of another physical process known as the “Auger effect.”¹

Under certain conditions of energy, an atom in a state of single ionization may undergo an *Auger transition* which leaves it in a state of double ionization. The electron released in such a transition is, in the final state, expelled from the atom and left with an amount of kinetic energy E given by

$$E = E_i - E_f, \quad (293)$$

where E_i is the initial energy of the singly ionized atom and E_f is the final energy of the doubly ionized atom. This type of transition does not involve the emission of a photon and, therefore, it is often called a “nonradiative” transition. Such transitions are possible, in general, only when an electron can be ejected from the atom; and for this to be possible it is necessary that E as defined by Eq. (293) be equal to or greater than zero. It is important to note that E is one value of a continuum of possible values.²

¹ COSTER and KRONIG, *Physica*, vol. 2, p. 13 (1935).

² For the benefit of students who may later study wave mechanics, it may be remarked that atomic states from which an Auger transition is possible do not represent true stationary states, or quantum states, of the atomic system. The atom in such a state cannot be represented by a Ψ function of the type of that written in Sec. 127 but must be represented by a Ψ or wave packet (Sec. 120) of a slightly more general form. As time goes on, Ψ changes, slowly or rapidly, into a form representing an additional electron of the atom as free and the others as remaining in the atom, which is in a new state of higher ionization.

Electrons ejected by such radiationless transitions were first observed by Auger, by means of the short, fat tracks made by them in the gas of a cloud chamber (Sec. 213) irradiated with X-rays.¹

A great many possible Auger transitions may occur satisfying Eq. (293), starting from a given initial state, and, furthermore, many different initial states are possible. To illustrate a particular transition, suppose E_i refers to the L_I state of an atom of atomic number Z , i.e., E_i is the energy of the L_I absorption limit. If the final state for an Auger transition is $L_{III}M_{IV}$, the value of E_f is the sum of the energies required to remove, first, an electron from the L_{III} shell and, then, a second electron from the M_{IV} shell. Of these two energies, the first is that corresponding to the L_{III} absorption limit of the atom of atomic number Z ; the second is practically equal to the energy corresponding to the M_{IV} absorption limit of an atom of atomic number $Z + 1$, since the absence of the L_{III} electron will cause the atomic field acting on the M_{IV} shell of electrons to resemble more closely that in an atom of nuclear charge increased by unity. Hence, as a good approximation, we may write as the energy equation for the transition $L_I \rightarrow L_{III}M_{IV}$

$$E = (E_{L_I})_Z - (E_{L_{III}})_Z - (E_{M_{IV}})_{Z+1}. \quad (294)$$

Clearly, nonradiative or Auger transitions provide (1) a means by which an atom can leave a given atomic state without emitting a photon, and (2) a means whereby atoms singly ionized by cathode-ray bombardment may automatically become doubly ionized. The Auger transition described in the preceding paragraph would (1) weaken the L lines that start from the L_I atomic state by decreasing the number of atoms in that state, and (2) increase the intensity of the accompanying satellite lines that start from the $L_{III}M_{IV}$ states.

In order that the Auger transition $L_I \rightarrow L_{III}M_{IV}$ may occur, the right hand member of Eq. (294) must be zero or positive, for E cannot be negative. In other words, we must have

$$(E_{L_I} - E_{L_{III}})_Z \geq (E_{M_{IV}})_{Z+1}. \quad (295)$$

The range of elements for which this condition is satisfied can be determined from a table² of the energies. In Fig. 178 are plotted values of the two members of Eq. (295) and of the similar equation for the transition $L_I \rightarrow L_{III}M_V$, as functions of Z . It is evident from the figure that the Auger transitions in question are possible for $Z < 50$ and for $Z > 75$ (about). These are just the regions of Z for which the

¹ AUGER, *Comptes Rendus*, vol. 180, p. 65 (1925).

² SIEGBAHN, *loc. cit.*

intensity of the $L\alpha$ satellites is anomalous. Hence, if we suppose that few atoms in the states $L_{III}M_{IV,V}$ are produced directly by the cathode-ray bombardment, the peculiar behavior of the $L\alpha$ satellites as to intensity is explained.

In a similar way, the Auger effect involving other nonradiative transitions has served to explain other anomalies of satellite intensity.

The second effect anticipated above is that lines for which L_I is the initial state should be weakened by the Auger transition just discussed and, hence, should undergo a rather abrupt change in

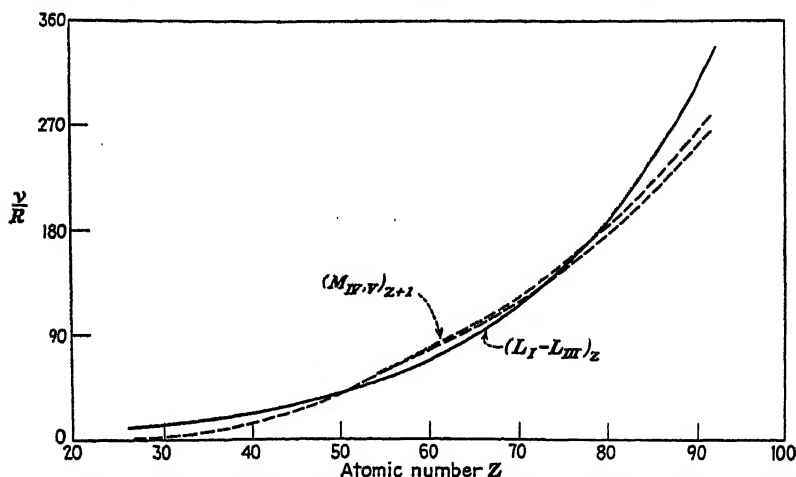


FIG. 178.—Coster-Kronig diagram for the L satellites; curves showing ν/R value of the $L_I - L_{III}$ energy-difference for an atom of atomic number Z , in comparison with the binding energies of M_{IV} and M_V electrons in atoms of atomic number $Z + 1$.

intensity at $Z = 50$ and $Z = 75$ (about). This effect, also, has been observed experimentally, as well as similar Auger effects on other lines.¹

Finally, it may be mentioned that the Auger effect furnishes an explanation, which cannot be discussed further here, of the great broadness of all lines except those of the K series.²

202. X-ray Spectra and the Outer Part of the Atom.—The usual emphasis in the study of X-ray spectra is on the elucidation of *inner* atomic structure. The spectroscopic theory discussed in this and preceding chapters of this book, however, was developed for gaseous atoms, whereas atoms emitting X-rays are generally packed together in a solid material. The spectroscopic theory developed for free

¹ COOPER, J. N., Cornell Dissertation; *Phys. Rev.*, vol. 59, p. 473 (1941).

² Cf. RICHTEMYER, BARNES, and RAMBERG, *Phys. Rev.*, vol. 46, p. 836 (1934); PARRATT, *Phys. Rev.*, vol. 54, p. 99 (1938).

atoms could be utilized in dealing with X-ray spectra because of the convenient fact that levels arising solely from a vacancy in one of the *innermost* shells of the atom are little disturbed by interactions with neighboring atoms in a solid. Any levels that involve changes in the *outer* part of the atom, however, will be greatly disturbed by this interaction and, hence, will be very different in a gas and in a solid.

In a few instances, X-ray spectra involving gaseous atoms have been studied. Emission spectra from such atoms have not been extended to include transitions involving changes in the outer part of the atom because of the extremely low intensities, but several absorption spectra of this type have been observed.

In our discussion of the absorption of X-rays in Sec. 190(b) and elsewhere, it was always assumed that, when an X-ray photon is absorbed, an electron is removed entirely from an atom. It should be possible, however, with proper incident $h\nu$, for the electron to stop in some outer vacancy in the atom. In this case the absorbed energy $h\nu$ would be less than if the electron were removed to infinity.

Wave mechanics furnishes the following basis for such transitions. With an electron removed from an inner shell, the ion will be surrounded by a field, due to the nucleus and the remaining electrons, which will approximate that of a hydrogen nucleus. For an electron moving in this field, there will exist a discrete set of quantum states and associated wave functions. The field can be regarded as the extension outside of the atom of the equivalent central field that is introduced in the zero-order stage of perturbation theory (Sec. 128). The electrons remaining in the atom can be regarded as occupying the innermost electronic quantum states in this field, leaving all the rest of them unoccupied. It should, therefore, be possible for the electron removed from an inner shell, instead of proceeding to an indefinite distance, to stop in one of these unoccupied outer states, *i.e.*, to change from an *inner* wave function belonging to an inner subshell to an *unoccupied outer* wave function.

According to this theory, we are led to expect that, instead of a single level, such as the *K* level, as hitherto described, there would be a closely spaced *sequence of energy levels*, as suggested in Fig. 179. The uppermost of the levels in the figure, corresponding to the ionized atom, represents removal of the electron into a state of rest at infinity.

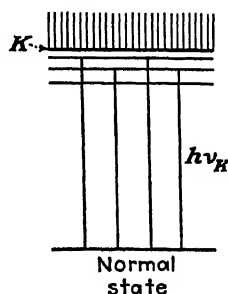


FIG. 179.—Illustration of a *K* photoelectric absorption limit and associated resonance levels.

The frequency ν_K corresponding to such removal of an electron from the K shell might be called the *photoelectric absorption limit for a gaseous absorber*. The lower-lying levels, corresponding to states in which the electron removed from the K shell remains attached to the atom and represented by one of the outer, previously unoccupied wave functions, are states of the neutral atom (sometimes called resonance levels).

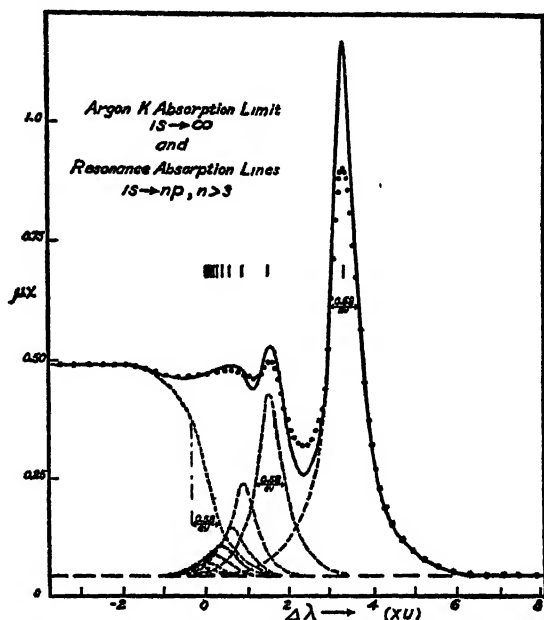


FIG. 180.—The K absorption edge of argon and its interpretation in terms of a series of resonance lines and the continuous photoelectric K absorption band. (Illustration furnished by courtesy of Prof. L. G. Parratt.)

A graph of the absorption coefficient should show, therefore, the well-known photoelectric absorption extending toward higher frequencies from the K photoelectric absorption edge and, in addition, on the long-wave side of this edge, a series of absorption lines, each arising from atomic transitions into one of the lower-lying atomic levels just described; the absorption edge would constitute the series limit for these lines. The lines should be very closely spaced and observable only with a spectrometer of very high resolving power; for the energy differences between the levels should be of the same order of magnitude as the differences between the ordinary optical levels, *i.e.*, a few electron-volts or less. The levels associated with a vacancy in one of the inner shells of an atom of atomic number Z ought, in fact, to correspond almost exactly, in relative position, to the ordinary optical

levels for a neutral atom of atomic number $Z + 1$; for the number of electrons in the outer part of the atom is the same in the two cases, and their motion should be affected by the removal of an inner (negative) electron in almost the same way as by an increase of the (positive) nuclear charge by 1 unit. A similar structure is to be expected at all of the absorption limits.

In Fig. 180 is shown an absorption curve for argon near its K limit,¹ as observed with a spectrometer of very high resolving power. There is clear evidence of the type of resonance absorption structure just described, but the absorption lines are so broad that there is great overlapping. The electronic configuration for argon ($Z = 18$), $1s^2 2s^2 2p^6 3s^2 3p^6$, will be altered in absorption to $1s^2 2s^2 2p^6 3s^2 3p^6 np$, a p electronic wave function being the only kind to which an electron can change from a $1s$ function in accordance with the selection rules. The spacing of the atomic levels under discussion will thus be almost the same as that of the optical P levels of potassium ($Z = 19$). Hence, if we ascribe the most intense line in Fig. 180 to the electronic transition $1s \rightarrow 4p$, we can locate the positions of the other resonance lines and of their series limit by using the known optical P terms for potassium. The photoelectric absorption edge, drawn in the figure to agree in wave length with the series limit, should have a finite "width" when observed with such high resolving power. Theory indicates that the shape of an absorption edge should be given by an arc-tangent curve, whose width is due to the same factors that cause the width of each of the absorption lines and also of the emission lines.²

203. X-ray Spectroscopy of Solids.—As mentioned in the previous section, the arrangement of the electrons in the outer part of an atom must be considerably altered when the atom is forced into close proximity with many other atoms, as in a solid. Consequently, we may expect the observed structure of absorption edges for atoms in a solid to be materially different from that found for a gaseous absorber, but, nevertheless, some type of resonance absorption is to be expected. Just as an interpretation of a curve such as that in Fig. 180 may provide us with information about the structure of a free atom, so may an interpretation of an absorption curve for a solid absorber reveal information about the basic structure of a solid. Furthermore, with a solid target in the X-ray tube, we may study the emission spectra involving the outer part of solid atoms and thus supplement the

¹ PARRATT, *Phys. Rev.*, vol. 56, p. 295 (1935). Resonance absorption structure, also found for polyatomic gases, was first identified by Kossel, *Zeits. f. Physik*, vol. 1, p. 119 (1920).

² See reference no. 2 on p. 550.

information obtained from the absorption curve. Based on experiments and studies of this sort, with such help as can be derived from an application of wave-mechanical theory, a *solid-state spectroscopy* is being developed.

A typical absorption curve for a solid absorber, showing the *K* region for copper, is reproduced¹ in Fig. 181. The curve shows obvious indications of resonance absorption, but no distinct lines stand out. Theoretical considerations indicate that an electron removed from the interior of an atom must be transferred to an electronic wave function that represents it as belonging to the entire solid rather than to an individual atom, and the energies associated with such transitions are distributed over broad bands of energy instead of being confined to a discrete set of values. The theory of solid-state spectroscopy has not developed, however, to the point that we can make a clear interpretation of the details of the observed structure of the absorption spectrum. For example, we cannot locate unambiguously the wave length of the photoelectric absorption edge for a solid absorber.

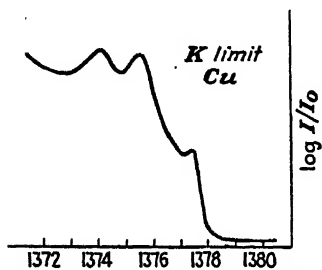


FIG. 181.—The *K* absorption edge for solid Cu (29). (I/I_0 = ratio of transmitted to incident intensity for a given thickness of copper; abscissa = wave length in X.U.)

The most fruitful work in emission in solid state spectroscopy has been in the long wave length region,² from 5 to 200 Å. Spectrographs using a mica crystal were employed to about 20 Å and a grazing-incidence ruled grating for $\lambda > 20$ Å. In such work, one studies, in particular, the X-ray emission lines that involve transitions of the atomic valence electrons, which have now come to belong to the solid as a whole. The effect of the interaction with neighboring atoms is equivalent to replacing the sharp levels of the free atom by bands of various width. The broadening of levels associated with vacancies deep in the interior of the atom, such as the *K* levels in the heavier atoms, is very slight. At the opposite extreme, however, levels associated with changes in the outermost atomic electrons are converted into broad bands of allowed energies. Thus the higher lines (short wave lengths) of the series of X-ray lines, as observed from the free atom, become replaced in the solid by broad emission bands, which

¹ BEEMAN and FRIEDMAN, *Phys. Rev.*, vol. 56, p. 392 (1939).

² SKINNER, *Phys. Soc., Reports*, vol. 5, p. 257 (1939), and references. See, however, reports by J. A. Bearden and collaborators, *Phys. Rev.*, vol. 58, pp. 387, 396, 400 (1940).

may overlap. These bands are emitted as valence electrons associated with the solid as a whole drop into vacancies produced by cathode-ray bombardment in the interior of atoms. The shape of these emission bands will be characteristic of the structure of the solid and may furnish information in regard to that structure.

Lack of space prevents an adequate treatment of this interesting development in X-ray spectra, but, to cite one conclusion, the Fermi-Dirac energy distribution of the valence electrons in a metal (Sec. 56) seems to be confirmed by recent studies of the shape of the X-ray bands emitted in the "valence $\rightarrow L$ " electronic transitions for aluminum.¹

¹ CADY and TOMBOULIAN, *Phys. Rev.*, vol. 59, p. 381 (1941) and references there given.

CHAPTER XI

THE NUCLEUS

In Chap. VI, we referred to Rutherford's experiments on the scattering of α particles and to the hypothesis of the nuclear type of atom which these experiments suggested. This was the first evidence for the existence of nuclei within atoms. The subsequent developments of quantum theory and its use in explaining the origin of spectral lines confirmed Rutherford's hypothesis and gave some information concerning the mass and the charge of the nucleus. Practically the entire mass of the atom is contained in the nucleus, and its charge is equal to Ze , where Z is the atomic number of the atom and e is the numerical value of the charge on an electron. The data furnished by ordinary spectroscopy and by chemistry yielded little additional information. More recently, however, new lines of attack have been developed, and at the present time (1940) our knowledge of the nucleus is well advanced and is still growing rapidly. Progress has resulted from work in several fields: (1) radioactivity, natural and artificial; (2) the precision measurements of the masses of atoms by means of the mass spectrograph; (3) artificial transformations or transmutations of nuclei by bombardment with (a) particles from radioactive substances, or (b) high-speed protons or other charged particles produced by laboratory methods, or (c) neutrons, or (d) γ -rays; (4) spectroscopic investigations in the visible and the ultraviolet regions, which yield evidence regarding (a) the angular momentum or "spin" of the nucleus and its magnetic moment, (b) the relative masses of certain isotopes and, in a few cases, (c) the existence of certain isotopes, the presence of which had escaped detection by the mass spectrograph; (5) the direct measurement of nuclear magnetic moments by means of the molecular-ray methods devised by Rabi (Sec. 162). In this chapter, we shall give a brief survey of some of these fields and of the information obtained concerning the structure and the properties of atomic nuclei.

THE MASSES OF ATOMS

204. Positive Rays.—Measurements of the masses of atoms are made by observing the deflection produced on positively charged ions of the substance under study by the combined action of an electric and a magnetic field, the methods used being similar to that by which J. J.

Thomson first measured the value of e/m for electrons. These positive ions in motion are frequently called *positive rays*. Three sources of positive rays are customarily employed: (1) the canal rays (see Sec. 161) originating near the cathode of a tube containing gas at a low pressure through which an electrical discharge is passing; (2) the positive ions emitted by salts when heated¹ under certain conditions; and (3) positive ions resulting from the vaporization in the discharge tube of the substance under study. The first source is used in studying those substances which can be conveniently introduced into the discharge tube in gaseous or vapor form; the second and the third when the substance is available only in the solid state.

The apparatus employed by J. J. Thomson for the study of positive rays² is shown diagrammatically in Fig. 182. B is a large discharge

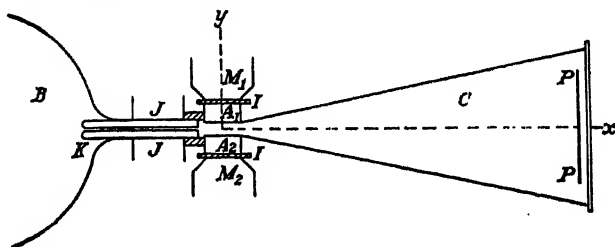


FIG. 182.—Thomson's positive-ray spectrograph.

tube, the cathode K of which is perforated with a very small hole through which pass the positively charged particles originating in the region immediately in front of K . These particles emerge from the opposite end of K as a narrow bundle of canal rays, the velocity of which depends on their charge and mass and on the potential V applied to the tube. JJ is a waterjacket for cooling the cathode. M_1 and M_2 are the poles of an electromagnet, of which A_1 and A_2 are the soft iron pole pieces electrically insulated from M_1 and M_2 by insulating strips I . Thus A_1 and A_2 can be maintained at any desired potential difference, so that in the space between A_1 and A_2 we may have a magnetic field and, parallel thereto, an electric field. A positively charged particle moving from left to right in this space will experience a deflection in the plane of the paper, due to the electrostatic field, and at right angles to the plane of the paper, due to the magnetic field. After leaving the space between A_1 and A_2 , the deflected particle moves in the field-free, evacuated space inside the

¹ See REIMANN, A. L., "Thermionic Emission," 1934, Chap. VI.

² See THOMSON, J. J., "Positive Rays of Electricity," 2d ed., 1921; ASTON, F. W., "Mass-spectra and Isotopes," 1933.

"camera" C and falls on the photographic plate PP . Thomson found that the plate, after development, showed a series of parabolas.

Let a particle, as it enters the space A_1A_2 , be moving with velocity v , initially parallel to and coincident with the x -axis of a system of rectangular coordinates with origin in the space A_1A_2 . Let the mass of the particle be M and its charge E . Also, let F and H be, respectively, the electrostatic and the magnetic fields, and S the length of the path through the fields, it being assumed that the fields are constant over the length S and that they terminate sharply. As a result of the passage through the field during a time S/v , the particle will experience a deflection y' (from the x -axis), due to the electric field, given by

$$y' = \frac{1}{2} \frac{FE}{M} \left(\frac{S}{v} \right)^2,$$

and a similar deflection z' in the z direction due to the magnetic field, assuming that the deflection is small compared with the radius of the circular path, given by

$$z' = \frac{1}{2} \frac{HEv}{M} \left(\frac{S}{v} \right)^2.$$

(Electromagnetic units are assumed throughout.) After leaving the fields, the particle moves in a straight line. If the distance of the photographic plate PP from the space A_1A_2 is large compared with the length of the path S , the point where the particle strikes the plate will have coordinates y and z which are, respectively, proportional to y' and z' . The relation between y and z will be, therefore, the same as the relation between y' and z' that is obtained by eliminating v from the last two equations, *viz.*,

$$z^2 = C \frac{E}{M} \frac{H^2}{F} y,$$

where C is a constant depending on the dimensions of the apparatus. This is the equation of a parabola. Accordingly, particles having various velocities v as they enter the field space A_1A_2 , but having the same ratio E/M , should make a parabolic trace on the plate, as Thomson found. For the same value of the fields H and F , particles with *different* ratios of E/M should produce on the plate different parabolas. The fact that the traces found by Thomson were reasonably sharp indicated that atoms of a given kind all have the same mass.

By a systematic study of the relative positions of the various parabolic traces appearing on a series of plates, Thomson was able to

determine the origin of the traces. Traces due to H^+ , H_2^+ , O^+ , O_2^+ , CO^+ , etc., were identified. Knowing the masses of these "standards," the masses of atoms producing other traces could be determined. Thomson found that when neon, atomic weight 20.18, was introduced into the discharge tube, instead of a *single* trace, there were *two* traces corresponding to atomic weights 20.0 and 22.0. This suggested that there are two kinds of neon atoms, one having atomic weight 20.0 and the other having atomic weight 22.0. A direct method was thus made available for measuring the masses of atoms. This method was rendered more precise by Aston, whose "mass spectrograph" has very greatly extended our knowledge of atoms.

205. The Mass Spectrograph. (a) *Aston's "Mass Spectrograph."* Aston's improvements¹ in the methods of positive-ray analysis consist (1) in securing greater dispersion and (2) in bringing all ions with a given ratio E/M to a *focus* (instead of spreading them out into a

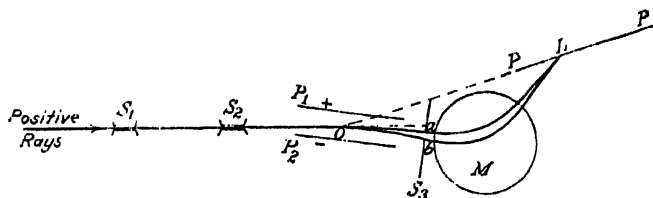


FIG. 183.—Diagram of the arrangement in Aston's "mass spectrograph."

parabola), thereby securing greater sensitivity. The principle employed is represented in Fig. 183. The positive rays from a discharge tube (not shown) pass through a narrow slit S_1 in the cathode, thence through a second slit S_2 , from which they emerge into the space between the metal plates P_1 and P_2 between which can be maintained an electric field of any desired intensity. This field causes a deflection of the particles toward P_2 , the amount of the deflection being greater the less the velocity of the particles. Since the positive-ray stream contains a wide range of velocities, the stream will be broadened as it passes through the field, the more swiftly moving ions passing through the wide slit S_3 on the side *a* and the more slowly moving ones passing on the side *b*. After leaving S_3 , this diverging stream of ions enters a magnetic field at right angles to the plane of the paper, maintained between the circular pole pieces M of an electromagnet. This magnetic field causes deflections, as shown, the more slowly moving ions being deflected more than the

¹ Aston's earlier papers are found in the *Phil. Mag.*, vols. 38–49 (1919–1925); see also *Nature*, vol. 116, p. 208 (1925); vol. 117, p. 893 (1926). An account of the subject is also given in Aston, *op. cit.*

faster ones, the result being that when the ions emerge from the magnetic field they are moving in *converging* directions in such a way that they are brought to a "focus" at some point L . It turns out that, with suitable design of the apparatus, the locus of the point L for ions having various ratios of E/M is nearly a straight line. Consequently, a photographic plate placed in position PP will record a number of "lines," each corresponding to a particular value of the ratio E/M . In part from the known dimensions of the apparatus, and in part by introducing into the discharge tube certain known substances— H_2 , O_2 , CO , CO_2 , etc.—as calibrating standards, the masses of other ions can be determined.

With this instrument Aston confirmed the existence of two kinds of neon atoms which differ from each other only in having *different masses*, these masses being very nearly the integers 20 and 22, respectively, when the mass of the oxygen atom is taken as 16.00. Similarly, when chlorine is introduced into the tube, no line is observed which corresponds to the chemically determined weight of chlorine, *viz.*, 35.46, but, instead, there are *two* lines corresponding very nearly to the integral values 35.0 and 37.0, the former being the more intense line. There are, thus, two kinds of chlorine atoms differing from each other in atomic weight but being *identical regarding all their other chemical and physical properties*. Hence, they are called *isotopes*.¹ Ordinary chlorine is a mixture of these two kinds of atoms, in such proportion that the average mass per atom of a large number of atoms—which is the mass given by chemical determinations—is 35.46.

In order to investigate whether the masses of atoms are *exactly* proportional to whole numbers as suggested in the preceding paragraph, Aston redesigned and refined his apparatus² so as to obtain a resolving power of 600 and an accuracy of the order of 1 in 10,000. It was then found that there are small, though very significant, departures from the whole-number rule. The masses of atoms are very nearly, *but not quite*, integers, taking the mass of the oxygen atom as 16. For example, whereas with his first mass spectrograph Aston found that the masses of the 2 isotopes of chlorine were 35.0 and 37.0, respectively, with the more accurate instrument these masses were found to be, respectively, 34.983 and 36.980. We shall return to this point presently.

(b) *Dempster's Method*.—Dempster's method of measuring the masses of atoms differs from Aston's in that (1) the photographic plate is replaced by an electrical method of measurement and (2) the

¹ From *isos*, equal; and *tópos*, place (in the periodic table).

² ASTON, *Roy. Soc., Proc.*, vol. 115, p. 487 (1927).

disposition of the electric and magnetic fields is materially different. His apparatus¹ is shown schematically in Fig. 184. An electrically heated metal cylinder *A*, which serves as anode, has on its front surface *f* a salt of the element under study. This heated salt is bombarded by electrons from the hot wire *w*, which is heated by the battery *B*₁, the wire being maintained at a potential difference *V*₂ of -30 to -60 volts with respect to *A* by means of the battery *B*₂. Dempster found that, when so bombarded, the anode emitted posi-

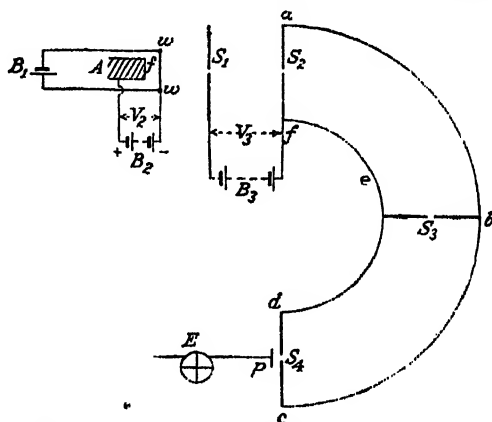


FIG. 184.—Dempster's "mass spectrograph."

tively charged ions. These ions, after passing through the slit *S*₁, are accelerated toward the slit *S*₂ by means of the potential difference *V*₃ of some 800 to 1,000 volts. On emerging from the slit *S*₂, the ions enter the space *abcdef* in which is maintained a uniform magnetic field, perpendicular to the plane of the paper, by means of which the ions are caused to move in a circular path toward the slit *S*₃, the radius *r* of the circle being determined by their velocity *v* as they enter *S*₂, their mass *M*, and charge *E*, as well as the intensity of the magnetic field *H*, according to the well-known equation

$$HEv = \frac{Mv^2}{r}.$$

Ions that move in circles defined by the three slits *S*₂, *S*₃, and *S*₄ pass through *S*₄ and fall upon the metal plate *p*, which thus acquires a positive charge at a rate which can be determined by the electrometer *E*. For a given ratio of *E*/*M* and field *H*, the value of *r* and, therefore, of the current *I* registered by the electrometer depends on the potential

¹ For Dempster's articles, see *Phys. Rev.*, vol. 11, p. 316 (1918); vol. 18, p. 415 (1921); vol. 20, p. 631 (1922).

difference V_s . A curve plotted between I and V_s shows sharp maxima, each of which corresponds to a definite value of E/M . In part, from the known constants of the apparatus and, in part, by use of known ions, Dempster was able to identify these maxima with definite ions.

When employing a salt containing lithium, Dempster found two maxima in the neighborhood of the atomic weight of lithium, 6.94, as shown in Fig. 185, in which the abscissas are the values of M , the computed mass of the ion, as determined semiempirically from the constants of the apparatus. The one maximum A corresponds within

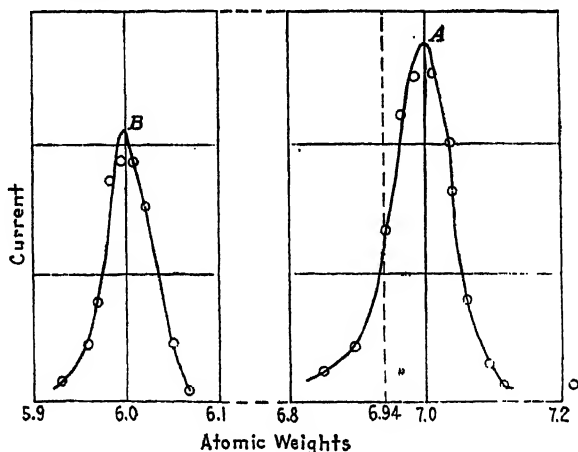


FIG. 185.—The isotopes of lithium, as measured by Dempster. The current scale for curve B is five times as large as for A .

the limits of error of measurement to atomic weight 7.00; the maximum B corresponds to 6.00. There is no indication of atoms of lithium with weight 6.94, the chemically determined atomic weight of lithium, which is indicated in Fig. 185 by the dotted line. By this method Dempster analyzed a number of elements—magnesium, lithium, calcium, and zinc—and found them to be made up of isotopes the masses of which are very nearly integers in terms of the mass of the oxygen atom as 16.

(c) *Bainbridge's Method*.—Recently, Bainbridge¹ has devised a mass spectrograph of high resolving power and precision, which has the great advantage of a linear mass scale. The apparatus is shown diagrammatically in Fig. 186. Positive ions, from a source not shown, enter through slit S_1 the space between the slits S_1 and S_2 . Between S_1 and S_2 is maintained a potential difference of several thousand volts. Ions of various velocities, masses, and charges pass through

¹ BAINBRIDGE, *Frank. Inst., J.*, vol. 215, p. 509 (1933).

slit S_2 into the "velocity selector" between S_2 and S_3 , in which there are crossed electric and magnetic fields, both perpendicular to the line joining S_2 and S_3 . The electric field is produced by maintaining the plates P_1 and P_2 at a fixed potential difference; the magnetic field, by an electromagnet, not shown. If F is the electric field and H the magnetic field, then only those ions will pass through S_3 which possess a velocity v given by

$$FE = HEv$$

or

$$v = \frac{F}{H}$$

where E is the charge on the ion. All other ions will be bent from the rectilinear path and will be lost from the beam. After passing S_3 , the ions enter a uniform magnetic field at right angles to the plane of the paper. In this field they travel in circles of radius R such that

$$HEv = \frac{Mv^2}{R}$$

or

$$R = \frac{v}{H} \frac{M}{E}$$

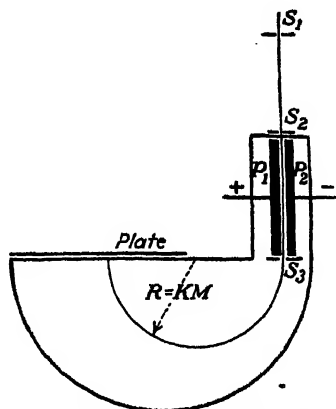


FIG. 186.—Bainbridge's mass spectrograph.

After traversing a semicircumference, the ions fall upon a photographic plate, leaving traces (lines), the position of which depends on the ratios M/E . It is readily seen that, for ions having a given charge E , M is proportional to R , and hence the mass scale is linear. Figure 187 shows the "mass spectrum" of germanium taken by Dr. Bainbridge.



FIG. 187. The mass spectrum of Ge(32), showing the isotopes 70, 72, 73, 74, 76. (From a print kindly furnished by Dr. Bainbridge.)

206. Isotopes. (a) *Isotopes of the Elements.*—All of the 92 elements have now been analyzed by means of the mass spectrograph. As has been implied in previous statements, the

atomic weights of the individual isotopes are commonly expressed on a scale on which the most abundant isotope of oxygen has a weight exactly equal to 16; such weights are called *isotopic weights*. They are found to be within 0.1 of an integer, which is called the *mass number* of the isotope in question, often designated by A .

The mass unit on the $O^{16} = 16$ scale, or the mass of an atom whose isotopic weight is unity, may be called the *atomic mass unit*, abbreviated as a.m.u. Its equivalent in grams is with sufficient approximation equal to the reciprocal of Avogadro's number (Sec. 45),¹ hence,

$$1 \text{ a.m.u.} = (6.023 \times 10^{23})^{-1} = 1.660 \times 10^{-24} \text{ gram.} \quad (296)$$

To indicate a particular isotope of an element, the mass number is commonly added to the chemical symbol as a superscript; the atomic number may also be added as a subscript, preferably preceding the symbol, thus:



Confusion is thus avoided with the ordinary use of the subscript in compounds; *e.g.*, the deuterium molecule is ${}_1\text{H}_2^2$, whereas ${}_1\text{H}^1{}_1\text{H}^2$ is a molecule containing 1 atom of ordinary hydrogen and 1 of deuterium (or heavy hydrogen, mass number 2).

A table of the known isotopes is given in Appendix I. Several of them were discovered spectroscopically (*e.g.*, H^2 , C^{13} , N^{15} , O^{17} , O^{18}). The relative abundance of the isotopic constituents of each element is shown in column 5. This can be determined either by comparing areas under such curves as those in Fig. 185 (allowing for any difference of scale) or from the intensities of the "line" produced on the photographic plate in the mass spectrograph.² Thus chlorine has been found to contain 75 per cent of Cl^{35} and 25 percent of Cl^{37} ; silver is 52.5 percent Ag^{107} and 47.5 percent Ag^{109} . Many isotopes are normally present only in very small quantities. For example, a recent study of calcium with the mass spectrograph has revealed the presence, in addition to masses 40, 42, 43, 44, and 48, of an isotope with $A = 46$, forming about 0.0033 percent of ordinary calcium.³

A glance at the table in Appendix I reveals a number of intriguing features in the distribution of isotopes as a function of atomic number. For some 17 of the 92 elements, only a single isotope is known (*e.g.*, Be, F, Na, Al, P, I, Cs, Au, Bi); and all but one of these elements (Be) have *odd atomic numbers*. For other elements, especially those of *even* atomic number, many isotopes are known, *e.g.*, as many as 10 for tin ($Z = 50$). Furthermore, isotopes of odd mass number tend to be both less numerous and less abundant in nature. Reasons for such features in the distribution of isotopes should be furnished by an

¹ Strictly speaking, (296) gives the weight of an atom of atomic weight unity [cf. Eq. (297) below], but there is no difference to the number of figures shown.

² For method, see Aston, *Roy. Soc., Proc.*, vol. 126, p. 511 (1930).

³ NIER, *Phys. Rev.*, vol. 53, p. 282 (1938).

adequate theory of nuclear structure. We shall return to this point later (Sec. 216).

(b) *Isotopic Constitution and Chemical "Atomic Weight."*—The atomic weight employed by the chemist obviously represents the average relative weight of the atoms in the mixture of isotopes with which he ordinarily works, or the *average isotopic weight* of this mixture. The atomic weight of oxygen is arbitrarily taken to be 16.

Oxygen, however, long thought to contain only the atom O^{16} , was later found to contain also minute amounts of O^{17} and O^{18} ; the masses of the latter, referred to O^{16} as 16 exactly, are, respectively, 17.0045 and 18.005. The average mass of the atoms in the normal oxygen mixture on the $O^{16} = 16$ scale is about 16.0044, as may be determined from the data given for oxygen in Appendix I. Thus, the atomic weight M_o for any isotopic mixture is related to M_i , the average isotopic weight of the atoms in the mixture, by the equation

$$M_o = \frac{16.0000}{16.0044} M_i = \left(1 - \frac{1}{3,600}\right) M_i. \quad (297)$$

As an example, ordinary lithium has been found to consist of 8.3 percent of Li^6 , with an isotopic weight $M_i = 6.0145$, and 91.7 percent of Li^7 , for which $M_i = 7.0146$; from these data, its average isotopic weight is readily found to be $M_i = 6.929$. The chemical "atomic weight" will be a little smaller, *viz.*, $M_o = 6.927$. This agrees sufficiently well with the value of 6.940 measured chemically (the ratios of abundance being known much less accurately than the mass numbers themselves).

In this way the chemical atomic weights of a number of the elements have been calculated from the measured relative abundance and masses of their respective isotopes.¹ The results are shown in column 6 of Appendix I. For comparison, the directly observed chemical atomic weights are shown in column seven. The agreement is seen to be excellent.

(c) *Separation of Isotopes.*—Chemically, the isotopes of a given element are almost or quite indistinguishable, since chemical properties depend primarily upon the nuclear charge and are little affected by the nuclear mass. The same statement must be true, in general, of physical properties, for these are mostly determined by the force fields surrounding the molecules, which depend in turn upon the number of electrons in the molecule and the electrical charges on the nuclei but

¹ See ASTON: *Roy. Soc., Proc.*, vol. 126, p. 511 (1930); vol. 130, p. 302 (1931); vol. 132, p. 487 (1931); HAHN, FLÜGGE, and MATTAUCH, *Phys. Zeits.*, vol. 41, p. 1 (1940).

are little influenced by the nuclear mass. One exception, of course, is the density, which must be nearly proportional to the isotopic weight. Any method for the separation of isotopes, therefore, must be based either directly on the difference in nuclear mass or on the utilization of very slight differences in other molecular properties.

The mass spectrograph constitutes the most obvious means for separating the isotopes of the elements. By replacing the photographic plate *PP* in Aston's apparatus (Fig. 183) by suitably disposed slits, we should be able to collect one kind of neon atoms, say, in one compartment and the other kind in another. Or, many kinds of atoms might simply be collected as deposits on a metal plate. In this way fairly pure samples of Li^6 and Li^7 have actually been obtained, also samples of uranium, U^{238} , free from U^{235} , up to 10^{-5} gram.

Many other methods of separating isotopes have been tried, based on slight differences in the rate of diffusion, of evaporation, of solution, and so on. Marked success has been attained, however, only in one case. In the case of hydrogen, electrolytic separation happened to be very efficient, and this method is employed on a large scale in the production of heavy hydrogen, H^2 . The isotope H^1 is evolved, in proportion to the amount present, five times¹ as fast as H^2 ; thus, if a large volume of water is almost but not quite completely electrolyzed, the small residue remaining is almost pure heavy water, H_2^2O .

RADIOACTIVITY

We have seen that the mass spectrograph yields very precise data concerning the masses of atoms and their nuclei. The fact that the masses of atoms are very nearly proportional to whole numbers raises two questions: whether the various nuclei may not be built up out of one or more common constituents; and whether or not it may be possible to transmute one nucleus into another. The first evidence suggesting an answer to these questions came from the field of radioactivity.

207. Becquerel's Discovery of Radioactivity.—The discovery of the phenomenon of radioactivity, although quite accidental, resulted directly from the discovery of X-rays. Roentgen had shown that X-rays are emitted by those parts of the discharge tube which are bombarded by the cathode rays. This bombardment was also accompanied by the emission of the well-known greenish or bluish fluorescence. The question arose: Is fluorescence always accompanied by the emission of X-rays?

¹ LEWIS and MACDONALD, *J. Chem. Phys.*, vol. 1, p. 34 (1933).

Several investigators had apparently found that fluorescent bodies activated by sunlight gave out a type of radiation which, like X-rays, was able to pass through black paper and to affect a photographic plate. In February, 1896, a few months after the discovery of X-rays, Henri Becquerel¹ was trying an experiment of this kind using as the fluorescing substance the double sulphate of uranium and potassium. After preparing the experiment and while waiting several days for sunshine, Becquerel discovered that even in the dark the specimen emitted a radiation which penetrated not only black paper but even thin sheets of metal, and that exposure of the fluorescing substance to sunlight had no effect on the phenomenon.

Becquerel soon found that this radiation was emitted by uranium irrespective of its state of chemical combination and that there was no connection whatever between this phenomenon and phosphorescence. Furthermore, the phenomenon was found to be quite independent of the temperature of the uranium compound. It was later discovered that these rays from uranium possess the power of discharging electroscopes by rendering the air through which they pass conducting.

This property of "radioactivity," as it was called, was soon found to be possessed by several other substances. Among them were thorium and two new elements, polonium and radium, discovered by M. and Mme. Curie, radium being more than a million times more active than uranium.

208. The Radioactive Radiations.—Like X-rays, the "rays" from radioactive materials affect a photographic plate, cause fluorescence, and ionize gases through which they pass. Unlike X-rays, however, these rays are of three types, which were given the names, before their nature was certainly known, of α -, β -, and γ -rays. If a small quantity of radium preparation is placed at the bottom of a small hole drilled into a lead block *B*, the emerging rays can be divided into the three groups by use of a strong magnetic field at right angles to the plane of the paper and directed away from the reader, as is shown in Fig. 188. One group is bent into a circular path to the right and will cause an impression on a photographic plate *PP*. These are the β -rays. From the direction of their deflection, it follows that they must be negatively charged particles. By studying quantitatively their deflection in

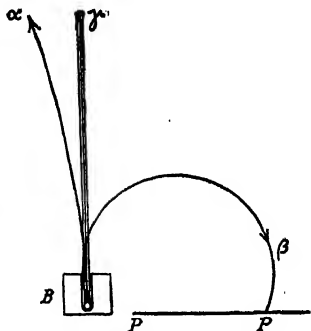


FIG. 188.—Schematic representation of the three types of rays from naturally radioactive materials.

¹ BECQUEREL'S papers appear in *Comptes Rendus*, vol. 122 (1896).

magnetic and electric fields, it was shown that these particles are electrons; they are ejected from radioactive materials with velocities which in some cases are very high. A second type of radiation is deflected slightly toward the left. This type consists of positively charged particles called " α particles," which were shown to possess a ratio E/M of the order of magnitude of that for *atoms*. These α particles were found to have a mass 4 (taking the mass of the oxygen atom as 16) and to carry a charge $+2e$. This identified them with the nuclei of helium atoms. The third type of radiation, the γ -rays, proceeds undeviated by either electric or magnetic fields, has a very high penetrating power, and is now known to consist of electromagnetic radiations of very short wave length lying, in general, in the spectral region beyond the shortest X-rays.

Recently it has been found that many artificially produced radioactive materials emit *positrons* instead of (negative) electrons, a positron being the same as an electron in all respects except that its electrical charge is positive (see Sec. 221). The tendency of usage seems to be to apply the term " β -rays" to both electrons and positrons when emitted by radioactive substances. In Fig. 188, the path of a positron β -ray would be a circle like that shown but curving toward the left instead of toward the right.

The three types of rays are further differentiated from each other by their penetrating power. The α -rays are absorbed by a few centimeters of air at ordinary pressure. They are reduced in intensity one half by passing through 0.005 mm. of aluminum. Their initial velocities are of the order of 2×10^9 cm.-sec.⁻¹. The β -rays are, roughly, 100 times more penetrating, since it requires something like 0.5 mm. of aluminum to reduce their intensity to half. The initial velocities of the β -rays, in some instances, exceed 99 percent of the velocity of light. The γ -rays are able to penetrate many centimeters of even so dense a metal as lead.

A short additional discussion of the properties of these radiations will be given in the next few sections. For further details, the reader is referred to special treatises¹ or to the original articles.

209. The Alpha-rays. (a) *General Properties.*—The velocity and the value of E/M for α -rays were determined by Rutherford by means of a modification² of the usual method in which magnetic and electric

¹ RUTHERFORD, CHADWICK, and ELLIS, "Radiations from Radioactive Substances," 1930; SODDY, "The Chemistry of the Radio Elements," 1914; HEVEY and PANETH, "Radioactivity," translated by R. W. LAWSON, 1936; F. RASSETTI, "Elements of Nuclear Physics," 1936.

² Cf. RUTHERFORD, CHADWICK, and ELLIS, *loc. cit.*

fields are employed. The ratio of charge to mass, E/M , was found to be almost exactly half as large for the α particles as it is for atoms of hydrogen. Additional observations were then necessary in order to determine E and M separately.

By means of the apparatus shown in Fig. 189, Rutherford and Geiger counted the number of α particles emitted per second by a given quantity of radium C.¹ The radioactive material was deposited on a disk D placed inside the highly evacuated vessel A at a known distance from a small circular opening S of known area, the opening being covered by a sheet of mica thin enough to allow the passage of the α particles into a brass chamber C . This chamber was evacuated to a pressure of several millimeters of mercury and had at its center an insulated wire WW , which, by means of a battery B , was maintained

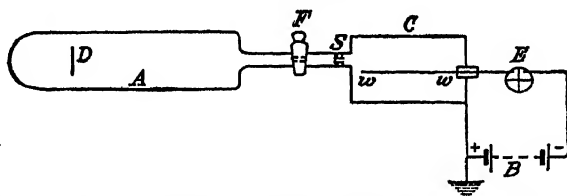


FIG. 189.—Apparatus of Rutherford and Geiger for "counting" α particles.

at a potential, with respect to the walls of the cylinder, just less than the critical discharge potential. When an α particle entered C , the ionization caused by its passage through the gas in C lowered the critical potential by an amount sufficient to allow the passage through the cylinder of a momentary current, which could be detected by a "kick" in the electrometer E . In this way, Rutherford and Geiger were able to detect the passage of *single* α particles. Thus, knowing the rate at which the α particles passed through S , the aperture which S subtended at D , and the quantity of radium on D , it was possible to determine the total number of α particles emitted per second per gram of radium C.

The total charge carried by the particles was then determined with the apparatus² shown schematically in Fig. 190. Through a window W , α particles from a known quantity of radium C deposited on the plate P were allowed to fall upon a collecting plate C ; the charge that they imparted to C was measured with an electrometer. To complete the electrical screen around C , the window was covered with aluminum foil thin enough to allow all of the α particles to pass through it. The entire vessel was evacuated.

¹ RUTHERFORD and GEIGER, *Roy. Soc. Proc.*, vol. 81, p. 141 (1908).

² *Ibid.*, p. 162.

Division of the observed total charge by the known number of the α particles then gave the charge on each one. This turned out to be twice as great as the electronic charge or the charge on a hydrogen ion. From this result and the previously found value of E/M , the mass of an α particle was found to be approximately four times that of a hydrogen atom. Thus the α particle must be the nucleus of a helium atom.

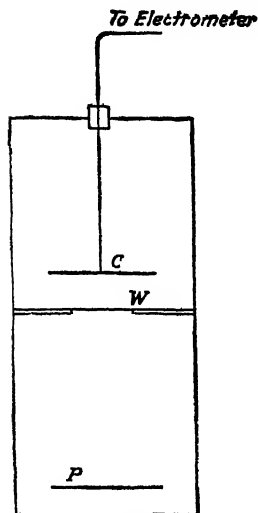


FIG. 190.—Rutherford's apparatus for measuring the charge carried by the α -rays from a known quantity of radioactive material.

This conclusion was confirmed, both qualitatively and quantitatively, by the observation that *helium* is produced as a result of α -ray activity. Rutherford's observations led to the conclusion that 1 gram of radium in radioactive equilibrium would emit 1.3×10^{11} α particles per second, and that these, after becoming neutralized by picking up electrons, would, in 1 year, form 0.16 cc. of helium gas.

The *velocity* of the α particles from radium C, determined as above by Rutherford, is 2.06×10^9 cm. per sec. Since their mass is

$$\frac{4}{N_0} = \frac{4}{6.023 \times 10^{23}} = 6.64 \times 10^{-24} \text{ gram,}$$

we find for their initial energy

$$\frac{1}{2}mv^2 = 14.1 \times 10^{-6} \text{ erg,}$$

or

$$\frac{14.1 \times 10^{-6}}{1.601 \times 10^{-12}} = 8.8 \times 10^6 \text{ electron-volts.}$$

More recent measurements give an initial velocity of 1.92×10^9 cm. per sec. and an initial energy of 7.68×10^6 electron-volts. It has been found that the α particles from a given substance may all have the same initial energy, or they may form several sets, the particles in each set having the same energy. In the latter case, when the α particles are separated by means of a magnetic field, a "line spectrum" is obtained.¹

In Table I below are shown initial velocities and energies for the α -rays from a number of typical radioactive substances.

(b) *The Range of Alpha Particles.*—Because of its double charge and great mass, an α particle produces an enormous number of ions as it passes through matter. Since energy is lost in producing these ions, and also in exciting many molecules without ionizing them, the α

¹ Cf. RASSETTI, *op. cit.*, p. 114.

particle rapidly loses velocity, until finally it is moving too slowly to produce any ions at all. The total length of path along which an α particle causes ionization is called its *range*. Since the range increases with increasing initial energy, it is commonly used as a measure of the initial energy of the α particle.

Various methods have been devised to measure the range of α particles. An approximate method makes use of the scintillations produced when α particles strike a screen of phosphorescent zinc sulfide, the range of the particles being measured by the maximum distance between source and screen, in air under standard conditions, at which scintillations can be observed. A more precise method is illustrated diagrammatically in Fig.

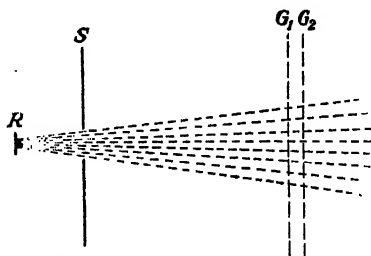


FIG. 191.—Arrangement for measuring the range of α particles from a source R by the ionization which they produce between two pieces of wire gauze, G_1G_2 .

191. The α particles from a radioactive source R pass through the opening in the screen S . The ionization produced in the space between two strips of wire gauze, G_1G_2 , 1 mm. or so apart, is measured for various distances between G_1G_2 and R (by varying either distance or gas pressure). Curves are obtained similar to that shown in Fig. 192 for radium

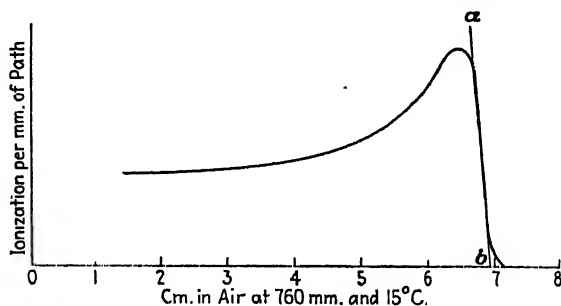


FIG. 192.—Ionization produced by α particles from radium C at different distances from the source. The ordinate represents about 25,000 pairs of ions per centimeter at the left end of the curve, or, at the maximum, about 60,000, in air of standard density.

C. The general shape of the curve may be accounted for as follows. Since the impulses given to the component parts of an atomic system by the passage near or through it of an α particle depend upon both the magnitude of the forces and the time during which those forces act, and since the speed of the particle is greatest at the beginning of its path, the smaller ionization over the first few centimeters is explained. As the particle loses energy, and therefore velocity, owing to the formation of ions, the times during which the forces act, and therefore the impulses,

increase for given distances of passage between α particle and atom, and the ionization per centimeter increases to a maximum; beyond this it falls rapidly. Since the exact end of the ionization curve is difficult to determine, it became customary to take the point b , where the straight line ab intersects the axis, as a measure of the range. A range so determined is called an *extrapolated range*.

The ordinate of the curve in Fig. 192 represents about 25,000 ion pairs per cm. in standard air at the left-hand end of the curve, and about 60,000 at the maximum of the curve. An ion pair consists of the ionized molecule and the electron ejected from it.

The ionization curves for α particles of lower initial energy are practically the same as a part of the curve shown in Fig. 192 taken from the proper point outward. For example, if the particles have such an initial energy that their extrapolated range is 4.9 cm., instead of 6.9 cm. as in Fig. 192, the curve will start at the point corresponding to $6.9 - 4.9 = 2.0$ cm. in the same figure.

Such curves are affected, however, by *straggling* of the particles; even if all of them have the same energy initially, their individual ranges differ slightly because of statistical fluctuations in the number and kinds of ions or excited molecules that they produce. By modifying the arrangement shown schematically in Fig. 191, so as to *count the number* of particles that cross the space G_1G_2 , instead of measuring the total ionization produced by them, the distribution of the individual ranges can be found. In the observations of Holloway and Livingston,¹ the counting was done by connecting G_1 or G_2 through an amplifier circuit to a thyatron, which, when suitably adjusted, has the property that voltage impulses below a certain magnitude have no effect upon it, whereas those exceeding this critical voltage cause a large momentary rush of current through the thyatron;² this current was made to activate a mechanical counter by means of an electromagnet.

Plotting the number of particles as thus observed against the distance x from R to G_1 (Fig. 191), we obtain a curve in which the ordinate represents the number of α particles that have traveled a distance equal to x without stopping. The slope of this curve then gives the number that stop per unit of x . In this way the true *mean range* of α particles of a given initial energy can be found. Doubtless in the future it will become customary to state as *the range* of the particles the mean range rather than the extrapolated range.

¹ HOLLOWAY and LIVINGSTON, *Phys. Rev.*, vol. 54, p. 18 (1938).

² See REICH, H. J., "Electron Tubes and Circuits," 1940; DOW, W. G., "Fundamentals of Engineering Electronics," 1937.

In experiments where it is not convenient to measure the range in air, it can be determined in metal foils whose "stopping power" in equivalent centimeters of air has been determined empirically. The range of the α particles from radium C in aluminum, copper, and lead is, in units of 10^{-3} cm., 4.06, 1.83, and 2.41, respectively. The "mass stopping power" of various elementary substances, or reciprocal of the range divided by the density, for α -rays, is nearly inversely proportional to the square root of the atomic weight. The (extrapolated) ranges in air of the α particles from a number of typical radioactive substances are given in Table I.

TABLE I.—RANGE, INITIAL VELOCITY, AND ENERGY OF ALPHA PARTICLES FROM TYPICAL RADIOACTIVE SUBSTANCES*

Emitter	Range, cm., in air at 760 mm. and 15°C.	Initial velocity, cm. sec. ⁻¹	Initial energy, ev.
Uranium I.....	2.73	1.40×10^9	4.05×10^8
Radium.....	3.39	1.51	4.74
Radium C.....	6.97	1.92	7.68
Radium F (polonium).....	3.93	1.59	5.25
Thorium.....	2.90	1.44	4.23
Thorium C'.....	8.62	2.06	8.76

* Compiled from a similar table given by Rutherford, Chadwick, and Ellis.

The increase in initial energy of the α particle that is required to lengthen its range by 1 cm. represents the loss of energy by the particle in going 1 cm. If we divide this loss of energy by the number of ion pairs produced in 1 cm., we have the average loss of energy per ion pair. This loss appears to be about 35 volts per ion pair, in air. Part of the energy, however, is undoubtedly lost in exciting molecules without ionizing them.

It may be remarked that *protons* (or hydrogen nuclei) produce about $\frac{1}{4}$ as many ions and lose about $\frac{1}{4}$ as much energy per centimeter as do α particles of the same velocity, owing to the fact that the ionization produced by a charged particle is roughly proportional to the square of its charge. Since the mass of the proton is likewise $\frac{1}{4}$ as great, its initial energy is also $\frac{1}{4}$ as great as that of an α particle moving with the same velocity. This means that protons and α particles of the same velocity have about the same range. Or, again, an α particle has 4 times the energy of a proton of the same range. If, however, we compare protons and α particles having the *same energy*, we find that the protons have ranges 5 to 10 times as great as the α

particles. A *deuteron*, or nucleus of heavy hydrogen (H^2), has twice as much energy as a proton of the same range.

210. Radioactive Transformations.—The early study of radioactivity led to the conclusion that the emission of α and β particles was associated with a chemical change in the radioactive atom. The transformation was found to follow an exponential law. During each successive element of the time, a certain definite fraction of the surviving atoms of the original substance “decay” or disintegrate, so that the total number surviving decreases in proportion to $e^{-\lambda t}$, where t is the time and λ is a constant, the decay constant. Thus, during a time $T = (\log 2)/\lambda$, half of the atoms disintegrate. The time T is called the *period*, or *half-value period*, of the radioactive substance. Instead of T , a time equal to $1/\lambda$ is sometimes cited; it is the *mean life* or average time during which an atom of the radioactive substance exists before disintegrating. [Cf. analogous relations in the emission of radiation, Sec. 125(a).]

The half-value period of uranium, for example, is estimated at 4.5×10^9 years, which is presumably long enough to account for the considerable quantities of uranium that exist in the earth's crust. In disintegrating, uranium gives rise, indirectly, to radium, with a period of 1,600 years; this period is so short, relatively, that radium is found only in uranium-bearing rocks, where it has been produced within the last few thousand years. Radium emits α particles and thereby becomes converted into an inert gas known as radon ($Z = 86$), or radium emanation; this gas also emits α particles, but it has a period of only 3.825 days.

If the new atoms formed as the result of radioactive change are themselves noticeably radioactive, many of their chemical properties, and the boiling point of the new element composed of them, can often be determined even if their mean life is so short that the new element cannot be accumulated in visible quantity. A radioactive substance can be followed in chemical reactions by means of its radioactivity, so that the chemist can tell, for example, whether it remains in solution or is precipitated or passes off as a gas. Usually a chemically similar substance, called a “carrier,” is added, so as to have something visible to work with; it is then easy to tell whether or not the radioactive substance follows the carrier in a chemical operation.

The nuclear theory of the atom made it possible to propose a precise theory of radioactive transformations which is in agreement with all of the known facts. When an atom, for example, loses an α particle, its mass number must decrease by 4, and its atomic number by 2, since the particle carries a double positive charge out of the nucleus. The

new substance formed will thus be an isotope of an element standing two columns to the left of the original substance in the periodic table. It may be one of the stable isotopes of that element which occurs in nature, or it may be a new isotope, stable or unstable.¹ That just one α particle is given out in such processes for each atom that is transformed has been shown by counting the number of α particles given out by a known amount of a radioactive substance in a given time and comparing the number so obtained with the known rate of disintegration of the material.

The ejection of a β particle from a nucleus, on the other hand, without changing its mass number, *increases* its positive charge and thereby *raises* its atomic number by unity. The emission of a γ -ray can obviously change neither the mass number nor the atomic number. It appears that γ -rays are not associated with an independent type of radioactive change but constitute an incidental accompaniment of transformations involving the emission of an α or β particle.

A systematic study of radioactive substances and their disintegration products has led to the discovery that the naturally radioactive substances can be grouped into several *radioactive series*, the elements of a given series being so arranged that each is a disintegration product of the preceding element. Thus, the uranium atom of mass number 238 ($Z = 92$), called "uranium I" or U I, may emit an α particle and thereby become converted into a nucleus with atomic number $Z = 90$ and mass number $A = 234$. The new substance so produced, called "uranium X_1 " (UX_1), is found by chemical tests to be an isotope of ordinary thorium ($Z = 90$, $A = 232$). The UX_1 atom then emits a β particle (period, 24.5 days) and becomes "uranium X_2 " (UX_2 ; $Z = 91$, $A = 234$); this emits another β particle (period, 1.14 minutes) and becomes uranium II (U II; $Z = 92$, $A = 234$). Thus U II is an isotope of ordinary uranium and is chemically indistinguishable from it; but its mass number is 234, not 238, and its period, a million years or so, is much shorter than that of U^{238} .

Uranium II also emits α particles (period, about 10^6 years), forming ionium, with $Z = 90$, $A = 230$. Ionium then emits another α particle (period, 7.6×10^4 years) and becomes radium ($Z = 88$, $A = 226$). The atomic weight of radium, determined chemically, is 225.97. The end product of this series, after the emission of five more α particles and four more β particles, is radium G or radium lead ($Z = 82$, $A = 206$), which is indistinguishable chemically from ordinary lead

¹ The transformed atom will promptly lose 2 of its circumnuclear electrons as well, thus bringing the number of its electrons into harmony with its reduced nuclear charge.

and has a measured atomic weight of 206.05. The entire uranium series is shown in Table II, and also, along with a similar series starting from thorium, in Fig. 193. In this figure, the emission of an α particle is indicated by an arrow pointing obliquely downward to the left, the emission of a β particle, by a short horizontal arrow pointing to the right. (Two small branches at UX_1 and at radium C are omitted.¹) Two radioactive members of the uranium series, Ra C and Ra E, are chemically indistinguishable from ordinary bismuth ($Z = 83$, $A = 209$).

TABLE II.—THE URANIUM SERIES OF RADIOACTIVE SUBSTANCES

Nucleus ..	UI	UX ₁	UX ₂	UII	Io	Ra	Rn
<i>Z</i>	92	90	91	92	90	88	86
<i>A</i>	238	234	234	234	230	226	222
Ray	α	β	β	α	α	α	α
Period...	4.5×10^9 years	24.5 days	1.14 min.	10^6 years	7.6×10^4 years		1,600 years

Nucleus ..	Rn	RaA	RaB	RaC	RaC'	RaD
<i>Z</i>	86	84	82	83	84	82
<i>A</i>	222	218	214	214	214	210
Ray	α	α	β	β	α	α
Period....	3.825 days	3.05 min.	26.8 min.	19.7 min.	(10^{-6} sec.)	

Nucleus ...	RaD	RaE	RaF	RaG (Pb)
<i>Z</i>	82	83	84	82
<i>A</i>	210	210	210	206
Ray.....	β	β	α	α
Period.....	(25 years)	5.0 days	136.3 days	

(U = uranium, Io = ionium, Ra = radium, Rn = radon, RaF = polonium.)

The end products of both the uranium and the thorium series are isotopes of lead, with $A = 206$ and 208, respectively. Ordinary lead has atomic weight 207.2; analysis by the mass spectrograph reveals the presence of isotopes 206, 207, and 208, with a small amount of 204. The atomic weight of lead, however, depends somewhat

¹ Cf. RUTHERFORD, CHADWICK, and ELLIS, *op. cit.*, Chap. I.

on the source. Uranium-bearing minerals usually contain lead—a fact in itself confirming the series of radioactive transformations from uranium to lead. The same is true of certain thorium-bearing minerals. The atomic weight of lead coming from uranium-bearing minerals is only a little over 206, whereas lead associated with Norway thorite has an atomic weight of 207.9, which is nearly the value 208 predicted for the end products of the thorium series.

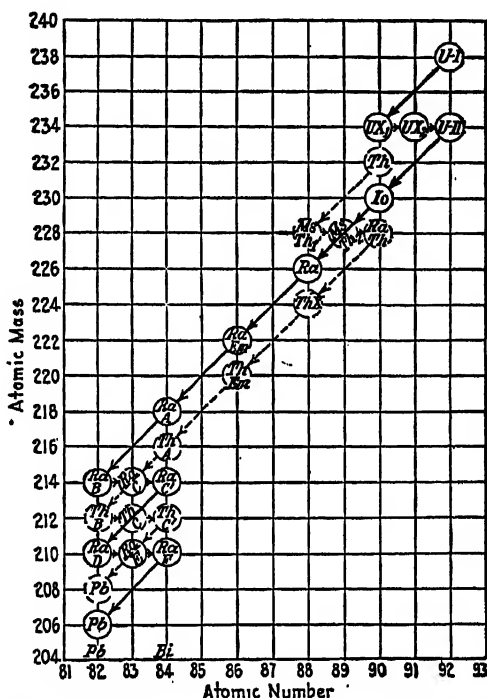


FIG. 193.—The radioactive-disintegration series of uranium and of thorium.

211. Gamma-ray Spectra and Nuclear Energy Levels.—Direct measurements of the wave lengths of some of the γ -rays have been made by use of the crystal grating. The wave length is so small, however, in comparison with the grating constants of available crystals that only the longer wave lengths can be measured in this way. Consequently, an indirect method has been developed based on the magnetic spectrum of the secondary β -rays which are produced by photoelectric absorption of the γ -rays in matter.

The method employed by Ellis¹ has been followed, with modification, by subsequent investigators. In Ellis's experiments the source

¹ ELLIS, *Roy. Soc., Proc.*, vol. 99, p. 261 (1921).

of the secondary rays was placed at W (Fig. 194) near a lead block B . W was a small glass tube 1 cm. long and 0.7 mm. in diameter, the inside walls of which were coated with a deposit of radium B which served as a source of the γ -rays under study. Around W was wrapped a foil of the metal M —Pb, Pt, W, etc.—which was to serve as the source of the secondary β -rays. The γ -rays passing out through the walls of the tube excited secondary β -rays in the metal foil. The whole apparatus was placed in a magnetic field (and, of course, in a

highly evacuated enclosure). The electrons passed, in circular paths, through the wide slit S and were "focused" onto the photographic plate PP at L . Upon developing the plate, a number of "lines" were found corresponding to various electron energies which could be determined by knowing the magnetic field and the diameter LW of the circular paths.

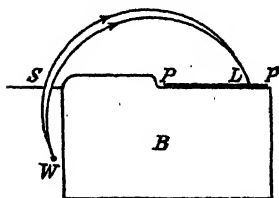


FIG. 194.—The β -ray spectrograph of Ellis.

The data in Table III, taken from Ellis's paper, will serve to illustrate the method of analyzing the results. For any given metal M , there was observed a number of lines on the plate of which three were more intense than the others. These same three appeared on the plate for each element but were shifted in position in such a direction as to indicate that the corresponding electrons had *less* energy the *greater* the atomic number of M . This is shown in the upper part of Table III, in which the energies of the electrons are expressed in electron-volts (Sec. 105). Ellis interpreted these electrons as *photoelectrons* ejected from certain levels in the atoms of the emitting substance. The change in the energy of the electrons in a given line from one element to another agrees closely with the difference of the energies of the K levels of the 2 atoms. Hence, he concluded that the electrons forming the three lines were expelled from the K levels of the respective atoms by three different γ -ray frequencies. Applying Einstein's photoelectric equation

$$h\nu = E_k + W_k,$$

where E_k is the observed kinetic energy of the photoelectron and W_k is the energy required to remove the electron from the K level of the atom, one obtains at once the energy $h\nu$ of the incident γ -rays. Data from each metal are found in this way to give the *same* group of three γ -rays, as the lower half of Table III shows. These, then, are three *lines* in the γ -ray spectrum of radium B, the respective wave lengths of which are 0.0518, 0.0425, and 0.0355 angstrom. In a

similar way energy values of many other lines in γ -ray spectra have been measured.

TABLE III.—DETERMINATION OF THE WAVE LENGTH OF GAMMA-RAYS FROM THE ENERGY OF SECONDARY BETA-RAYS (ELLIS)

Emitter—→	Tung- sten 74	Plati- num 78	Lead 82	Ura- nium 92		
Secondary β -ray energies in electron-volts $\times 10^{-5}$	1.66	1.58	1.49	1.22		
	2.20	2.12	2.03	1.74		
	2.76	2.69	2.60	2.31		
E_K of emitter—→	0.693	0.782	0.891	1.178	Mean energy	λ , angstroms
Energy ($h\nu$) of γ -rays in electron-volts $\times 10^{-5}$	2.35	2.36	2.38	2.40	2.37	0.0521
	2.89	2.91	2.92	2.92	2.91	0.0425
	3.46	3.46	3.49	3.48	3.47	0.0355

Photoelectrons from shells other than the K shell have also been observed. In addition, there is a continuous background in the spectrum of the secondary β -rays which is due to Compton electrons resulting from *scattering* of the γ -rays.

If the interpretation of such observations is correct, it will also be expected that a γ -ray may be absorbed photoelectrically by one of the electrons surrounding the nucleus that emits the γ -ray itself. Radioactive substances emitting γ -rays of a given wave length ought, therefore, to emit, also, groups of β -rays forming a spectrum of β -ray lines corresponding to ejection from the K , L , M , . . . shells of the radioactive atoms. Such is, indeed, observed to be the case. The production of *secondary* β -rays in this manner is commonly called *internal conversion* of the γ -ray. It happens much more frequently in the K shell than in the L shell; the probability that the γ -ray is so converted before escaping from the atom varies in different cases from 0.001 or less to 0.1. β -ray lines produced in this manner are much sharper (the velocities of the β -rays being more nearly uniform within each group) than are the lines excited by the same γ -rays in extraneous substances. Hence most wave lengths of γ -ray lines have actually been determined from observations of the energies of the internal-conversion β -rays.

Not all β -rays arise from internal conversion of γ -rays, however. As we shall see in the next section, many of them come from the nucleus itself.

All γ -ray spectra, apparently, consist of sharp lines. Some of them have wave lengths longer than the hardest X-rays obtainable from the same atom, but most of their wave lengths are shorter than the shortest K line from any element (*viz.*, 112 X.U. = 0.112 Å, from uranium). It seems clear, therefore, that the γ -rays cannot possibly consist of radiation emitted by the circumnuclear electrons; in view of the success of the modern theory of atomic spectra, we can assume with safety that the electrons can emit no harder lines than the K lines. The γ -rays must, therefore, be emitted somehow by the nucleus.

(a) *Nuclear Energy Levels.*—In some cases measurements of X-ray spectra have yielded sufficient data so that it has been possible to postulate a system of energy levels in the nucleus, transitions between

TABLE IV.—SOME LINES IN THE GAMMA-RAY SPECTRUM OF RADIUM B AND THE ENERGY LEVELS IN THE RADIUM C NUCLEUS (ELLIS AND SKINNER)

Line	Wave length, angstroms	Energy (unit: 10 ⁶ electron- volts)	Origin	Proposed energy levels (unit: 10 ⁶ electron-volts)
C_2^*	1.37	0.090	$B \rightarrow C$	$A = 0$
C_1	0.230	0.537	$A \rightarrow B$	$B = 0.537$
C_2^{**}	.196	0.629	$A \rightarrow C$	$C = 0.625$
C_3^*	.115	1.073	$E \rightarrow F$	$D = 2.572$
C_4^*	.098	1.250	$F \rightarrow G$	$E = 2.942$
E_1	.0634	1.947	$C \rightarrow D$	$F = 4.048$
E_2	.0607	2.035	$B \rightarrow D$	$G = 5.31$
E_3	.0513	2.404	$B \rightarrow E$	
E_4	.0480	2.572	$A \rightarrow D$	
E_5	.0451	2.733	$D \rightarrow G$	
E_6	.0419	2.942	$A \rightarrow E$	
E_7	.0351	3.511	$B \rightarrow F$	
E_{16}	.0263	4.684	$C \rightarrow G$	
E_{17}	.0257	4.800	$B \rightarrow G$	

* Measured by Rutherford and Andrade by crystal reflection.

which give rise to γ -ray lines in exactly the same way that X-ray lines originate in transitions between levels in the extranuclear structure of the atom.

The wave lengths and energy values of some of the lines in the γ -ray spectrum of radium B, as found by Ellis and Skinner,¹ together with the "name" of the line, are shown in the first three columns of Table IV. These investigators found it possible to postulate a series of seven levels A, B, C, . . . transitions among which give rise to the

¹ ELLIS and SKINNER, *Roy. Soc., Proc.*, vol. 105, pp. 165, 185 (1924).

14 observed lines, as indicated in column 4. The resulting energy-level diagram is shown in Fig. 195.

In view of the difficulty of making the measurements, it is a scientific achievement of the first magnitude to have isolated and measured a set of γ -ray lines with sufficient precision for the construction of an energy-level diagram. Like most scientific advances, however, this accomplishment raises fresh questions. First of all, how do the nuclei come to be in excited states out of which they can undergo transitions to lower levels? The situation is quite different

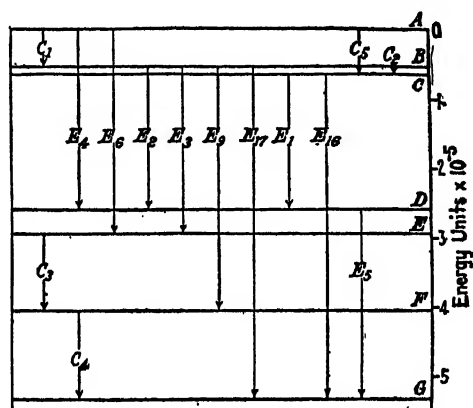


FIG. 195.—Energy-level diagram for the γ -rays from radium B. (ELLIS and SKINNER.)

from that in vacuum tubes, where atoms are continually being excited by electron impact.

A clue to the answer is furnished by the fact that (in natural radioactivity) γ -rays are emitted only by substances which also emit α - or β -rays. It is reasonable to suppose that the emission of an α or β particle leaves the residual nucleus not in its normal state but in an excited state of higher energy. If this is the correct explanation, then the γ -ray spectrum ascribed to radium B is really emitted by nuclei of the radioactive product, which is radium C; and the energy levels deduced from these rays are levels of the radium C nucleus, not of radium B. This conclusion has been confirmed by experiment, most clearly by Ellis and Wooster.¹ They compared the difference in the energies of the secondary β -rays produced by a given γ -ray line in the radioactive source and in platinum with the difference in the energies of the K levels in the two substances.

It appears that all natural γ -rays are emitted by nuclei which have just been formed as a result of α - or β -ray activity. The interval

¹ ELLIS and WOOSTER, *Cambridge Phil. Soc., Proc.*, vol. 22, p. 844 (1926).

of time between the formation of the new nucleus and the emission of the γ -ray is probably extremely short, of the order of 10^{-12} sec. Hence, the γ -ray activity is observed to decay at the same rate as the α - or β -ray activity to which it owes its origin, the rate of decay having reference primarily to the process of emission of the α or β particle.

Since the emission of a γ -ray thus involves no change in nuclear charge or mass number, being analogous in these respects to the emission of radiation by atoms, there is a strong tendency nowadays to restrict the terms "radioactive" and "radioactivity" to the emission of α or β particles.

212. Beta-ray Spectra.—When the β -rays from a radioactive substance are spread out into a spectrum according to their velocity, by means of a magnetic field, it is commonly found that the spectrum consists of narrow lines superposed upon a continuous background. In the beginning it was thought that all β -rays come from the nucleus, but later work appears to have shown, as we have indicated, that all of the *lines* are due to the ejection of *circumnuclear electrons* from the atom. In most cases these β -rays consist of photoelectrons produced by γ -rays. One line from a disintegration product of radium C, however, is perhaps due to direct interaction between an excited nucleus and the K shell of electrons, the nucleus jumping to a lower nuclear energy level without the emission of radiation and thereby furnishing the energy for the ejection of a K electron [in analogy with the Auger transitions in the X-ray region, Sec. 201(b)].

The β -rays forming the *continuum*, on the other hand, are believed always to come from the nucleus. In confirmation of this interpretation, measurements have shown that, in the emission of such *primary* β -rays, as they are called, just one ray is emitted for each nucleus that is transformed.

When the number of β -rays per unit of their energy is plotted against the energy as abscissa, a curve is obtained which rises quickly to a maximum and then sinks slowly to zero at an upper limit of energy. The observed maximum energies range from 6×10^5 to over 3×10^6 electron-volts. An experimental curve obtained for RaE (period, 5 days) is shown in Fig. 196.

Such curves show that the energy of the β -ray which is emitted by atoms undergoing a given type of radioactive transformation varies greatly from atom to atom. This fact presents great difficulties in regard to the conservation of energy. The kinetic energy of the β -ray should represent the energy lost by the nucleus in undergoing the transformation. The loss in energy might have one of several different values, according as the nucleus is left by the β transformation

in one or another of those nuclear levels whose existence is revealed by the γ -ray spectra; but a variation due to this cause would result in a line structure in the spectrum of the primary β -rays, instead of a continuum.

It was suggested at one time by Bohr that the law of the conservation of energy might not hold for the β -decay of nuclei. There is no other indication of a failure of this law, however. At the present time, preference is given to a hypothesis proposed by Fermi. He assumes that another particle of a new kind, called a *neutrino*, is emitted along with the β -ray. The neutrino is supposed to carry

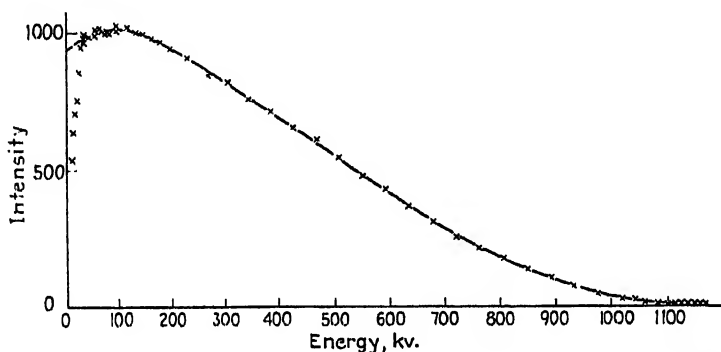


FIG. 196.— β -ray spectrum of RaE. The ordinate represents, on an arbitrary scale, the number of β -rays emitted per unit of their kinetic energy.

no charge and to have a very small mass, hence it can pass unhindered through miles of matter and so has eluded experimental detection. The *upper limit* of the β -ray energy is then regarded as representing the energy that is lost by the nucleus in its transformation; the difference between this energy and that of the β -ray is carried off by the neutrino. This theory, however, is still in the speculative stage.

213. Observations on Individual Charged Particles.—At this point it may be instructive to describe briefly some additional methods by which it has been found possible to make observations on *single* charged particles. The experiments described hitherto have been such as to yield only statistical results, from which the properties of individual particles could only be inferred by means of indirect reasoning.

In Sec. 209(a) above, we described the method by which Rutherford and Geiger counted the α -rays emitted from a radioactive substance. Each α particle was made to cause a separate pulse of ionization in an ionization chamber, and these pulses, detected with an electrometer, were counted. This method has been used extensively

to count particles of all sorts, under the name of a *Geiger counter*. Even γ -rays can be "counted" with such devices, the ionization in the counter being caused by secondary electrons ejected from molecules of the enclosed gas or from the walls of the counter.

Various changes have been made, however, in the ionization chamber, and still greater changes in the method of counting the impulses of ionization.¹ Nowadays the small electrical impulses from the ionization chamber are usually magnified many times by means of an electron-tube amplifier. In the experiments of Holloway and Livingston referred to in Sec. 209(b) above, amplified impulses up to 250 volts were observed, each due to the entrance of a single α particle into the ionization chamber.

For α particles, an alternative but somewhat less reliable method is to count the scintillations or flashes of light produced by them when they fall upon a phosphorescent screen, which may be coated with impure zinc sulfide. By comparing counts made in this way with those made using a Geiger counter, it has been shown that each scintillation is produced by a single particle. The individual scintillations can be seen easily under a microscope.

The most instructive method of studying individual charged particles, however, is to observe the tracks made by them in a *Wilson expansion chamber* or *cloud chamber*.² The method depends upon the fact that when dust-free air saturated with water vapor is expanded quickly and thereby cooled, if no ions are present, a considerable expansion is necessary in order to cause precipitation of the vapor, whereas if ions are present, a somewhat smaller expansion causes a droplet of water to condense about each ion as a center. In this way the path of an ionizing particle can be made visible as a row or "track" of cloud droplets condensed about the ions.

The cloud chamber is commonly made a few centimeters deep and considerably wider, and the expansion of the air contained in it is produced by jerking the floor of the chamber downward, *e.g.*, by suddenly releasing the air below it through a valve into an evacuated vessel. In a modification due to Shimizu, the floor of the chamber consists of a tight-fitting piston which is caused to move up and down periodically. Vapors of liquids other than water, such as alcohol, are often used. The particles to be observed are admitted to the

¹ Cf. RASETTI, *op. cit.*, p. 18; articles in *Rev. Sci. Instruments*, or articles describing observations made with the use of counters; WYNN-WILLIAMS, *Roy. Soc., Proc.*, vol. 136, p. 312 (1932).

² WILSON, C. T. R., *Roy. Soc., Proc.*, vol. 85, p. 285 (1911); vol. 87, p. 277 (1912); vol. 104, p. 1 (1923). Cf. RUTHERFORD, CHADWICK, and ELLIS, *op. cit.*, p. 57.

chamber just before the expansion; and, after the pressure has been restored, an electric field may be applied momentarily to clear the space of ions. The tracks are illuminated by a strong beam of light and are either observed visually or photographed through the glass top or sides of the chamber. By using two cameras pointing at different angles, stereoscopic pictures of the tracks are frequently made. (Cf. Fig. 223 in Sec. 234.) The whole apparatus is often arranged to operate automatically, expansions being produced at intervals of 10 to 30 seconds. Photographs may be taken on motion-picture film.

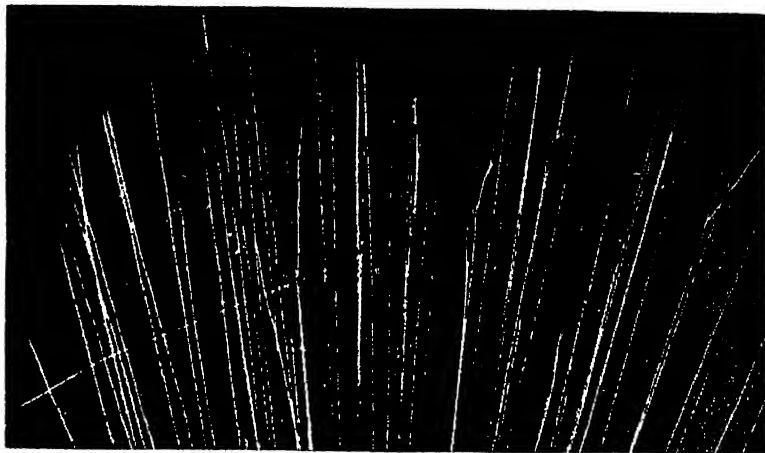


Fig. 197.—Cloud-chamber tracks of α particles from polonium.

Since an α particle produces something like 50,000 droplets per cm. in air at ordinary pressure, it makes a heavy, solid-looking track in a cloud chamber. The tracks are practically straight, except for an occasional kink where the α particle came very close to a nucleus and suffered an appreciable deflection (cf. Fig. 197). The distribution of such paths as to range is easily studied in a cloud chamber. Protons also make solid-looking tracks, but they are more slender. An electron, on the other hand, producing only perhaps 100 droplets per centimeter, looks thin and knobby; under a low-power microscope the individual droplets can be distinguished and counted (cf. Fig. 227 at Sec. 236). A γ -ray produces no track at all, except that starting from some point along its geometrical path there may be seen the short track of an electron ejected by it from a molecule of the gas.

The cloud chamber has been of immense service in the study of high-speed particles of all sorts, because of the possibility of observing

the effects caused by each individual particle. In the field of natural radioactivity, cloud-chamber observations have always yielded results in harmony with those obtained by other methods.

THE STRUCTURE OF NUCLEI

214. The Building Up of Nuclei. (a) *The Elementary Particles.*—

A survey of the data on radioactivity in the light of Rutherford's nuclear theory of the atom raises the question as to the constitution of the nucleus. A mechanism capable of emitting α -rays, β -rays, and a whole spectrum of monochromatic γ -rays must possess a well-ordered, though probably very complex, structure. And, first of all, of what elementary particles is the nucleus composed?

Since electrons were known to be emitted from some nuclei, it was natural to assume in the beginning that electrons formed one constituent of many, perhaps all nuclei. Such an assumption would make it possible to carry out Prout's old hypothesis in a new form, by assuming that all nuclei consisted of electrons and protons (a proton being the nucleus of a hydrogen atom). Radioactive nuclei have never been found to emit protons, to be sure; and they do often emit α particles, which are helium nuclei. But the helium nucleus, being almost exactly four times as heavy as a proton with a positive charge twice as great, might consist of 4 protons combined with 2 electrons, the latter adding very little to the mass but serving to make the nuclear charge come out right.

On the basis of wave mechanics, however, certain objections can be raised against the assumption that electrons exist inside of nuclei. According to the indeterminacy principle (Sec. 121), the indefiniteness of position of a particle in one direction, Δx , and the indefiniteness of the corresponding component of momentum, Δp , are always at least large enough so that

$$\Delta x \Delta p \cong h,$$

h being Planck's constant and \cong meaning "is of the order of." Now, if we make $\Delta x = 2 \times 10^{-12}$ cm., which is of the order of nuclear diameters [cf. Eq. (307) in Sec. 215],

$$\Delta p = \frac{h}{\Delta x} = \frac{6.6 \times 10^{-27}}{2 \times 10^{-12}} = 3.3 \times 10^{-15}.$$

To estimate the indefiniteness in energy, we must use relativistic formulas. For the total energy W of a particle with momentum p , Eq. (73) in Sec. 69 gives

$$W^2 = m^2 c^4 + p^2 c^2.$$

If we make p only as large as the value just found for Δp , the term $p^2c^2 = (3.3 \times 10^{-15})^2(3 \times 10^{10})^2 = 10^{-8}$ erg². For an electron, this is easily seen to exceed greatly the term m^2c^4 . Hence for such a value of p we may write

$$W^2 = p^2c^2 = 10^{-8}, \quad W = 10^{-4} \text{ erg} = \frac{10^{-4}}{1.6 \times 10^{-12}} = 6 \times 10^7 \text{ ev}$$

[cf. Eq. (144) in Sec. 105]. Only a small part of W is represented by the rest energy

$$mc^2 = (0.9 \times 10^{-27}) \frac{(3 \times 10^{10})^2}{1.6 \times 10^{-12}} = 5 \times 10^5 \text{ electron-volts.}$$

Thus a free electron confined within a space as small as a nucleus would have to have a kinetic energy of the order of 6×10^7 electron-volts. In a nucleus, no doubt, an electron would really be in some sort of quantum state; but a more rigorous wave-mechanical argument leads to the same conclusion as to its energy. Yet, experimentally, we find that when an electron issues from a nucleus as a β -ray, its kinetic energy never exceeds 4×10^6 electron-volts, or $1/15$ of that just found. In view of the contrast between these two numbers it seems improbable that nuclei can contain electrons. There are other theoretical objections connected with nuclear spin, but these we shall not discuss.

In view of these difficulties with the assumption that nuclei contain electrons, the discovery of the *neutron* in 1932 led at once to the general adoption of a radically different theory of nuclear composition. The properties of the neutron will be discussed in detail later (Secs. 220, 223). All that we need to know about them here is that they are particles having almost the same mass as a proton but no electrical charge at all. In the theory of nuclear composition generally held today it is assumed that

1. Nuclei are composed of protons and neutrons.
2. When an electron is emitted from a nucleus as a β -ray, it is created in some way at the instant of emission, a neutron in the nucleus changing at the same time into a proton. If a positron is emitted, a proton changes into a neutron.

Thus the idea of the conservation of electrons has been abandoned, but conservation of electrical charge is retained.

It will be instructive to survey the existing array of nuclei on the basis of this theory.

(b) *The Formation of Nuclei.*—According to the theory of nuclear composition just stated, the helium nucleus, with atomic number $Z = 2$ and mass number $A = 4$, must consist of 2 protons and 2 neutrons. But, when the mass of the helium nucleus is compared with the combined masses of 2 protons and 2 neutrons, an appreciable discrepancy is found to exist.

The mass of the neutral hydrogen atom, H^1 , in terms of atomic mass units ($O^{16} = 16$, cf. Sec. 206), as measured with the mass spectrograph, is¹

$$H^1 = 1.00812 \text{ a.m.u.}, \quad (298)$$

the last figure being uncertain. Here, for simplicity, we have let the atomic symbol stand also for the mass of an atom; no misunderstanding will result from this double use. Dividing this number by 1,837, the ratio of the electron mass m to H^1 (Sec. 45), we have m in mass units; and then, subtracting m from H^1 , we have the mass of a proton; this we shall indicate by the symbol p , which is often used to denote a proton. Thus we obtain

$$m = 0.00055 \text{ a.m.u.}; \quad p = 1.00757 \text{ a.m.u.} \quad (299a,b)$$

The mass of a neutron, on the other hand, is (Sec. 220)

$$n = 1.00893 \text{ a.m.u.}, \quad (300)$$

the last figure being uncertain. The mass of a neutral helium atom, on the other hand, as measured with the mass spectrograph, is

$$He^4 = 4.0039 \text{ a.m.u.} \quad (301)$$

If we subtract from the mass of neutral helium the combined masses of 2 neutral hydrogen atoms and 2 neutrons, the result is the same as the difference between the mass of a helium *nucleus* alone and the combined masses of 2 *protons* and 2 neutrons; the masses of the 2 circumnuclear electrons, which are included in the mass of the neutral helium atom, are canceled by the masses of the electrons in the 2 hydrogen atoms. (It is customary in nuclear calculations to work in this way with the masses of *neutral atoms*, for the reason that it is usually these that are cited as the result of experimental measurement.) Since, from the numbers just cited,

$$2H^1 + 2n = 4.0341,$$

¹ BARKAS, *Phys. Rev.*, vol. 55, p. 691 (1939). All atomic masses quoted up to $Z = 40$ will be taken from this article.

we find that, if 2 protons and 2 neutrons were to combine into a helium nucleus, there would be a loss of mass of

$$\Delta M = 4.0340 - 4.0039 = 0.0300 \text{ a.m.u.} \quad (302)$$

A plausible interpretation of this loss of mass suggests itself when we recall the relativistic relation between mass and energy, described in Sec. 69, according to which each erg of energy is accompanied by $1/c^2$ grams of mass, c being the speed of light.¹ The loss of mass would thus be accounted for if the binding of the protons and neutrons into the helium nucleus were accompanied by a corresponding loss of energy.

The amount of energy that is associated with an atomic unit of mass is c^2 times the unit of mass, whose magnitude is given in Eq. (296) in Sec. 206. We find thus that 1 atomic mass unit or

$$1 \text{ a.m.u.} = 1.660 \times 10^{-24} \times (2.998 \times 10^{10})^2 = 1.492 \times 10^{-3} \text{ erg};$$

or, if we divide by 1.601×10^{-12} erg [Eq. (144) in Sec. 105], we find that $1 \text{ a.m.u.} = 9.32 \times 10^8$ electron-volts. It is customary in nuclear work to use ev (or eV, EV) as an abbreviation for electron-volts and Mev (sometimes MEV, MV) for a million electron-volts. Thus

$$1 \text{ a.m.u.} = 1.492 \times 10^{-3} \text{ erg} = 932 \text{ Mev.} \quad (303)$$

For future use, we may note also the energy equivalent of the electronic mass. Calculating m as in obtaining Eq. (299a) and using Eq. (303), we find $mc^2 = 0.000549 \times 932$ or

$$mc^2 = 0.511 \text{ Mev.} \quad (304)$$

Combining Eqs. (302) and (303), we have then, as the loss of energy in the formation of a helium nucleus, or its *binding energy* relative to free protons and neutrons,

$$\Delta W = 0.0300 \times 932 = 28.0 \text{ Mev.} \quad (305)$$

This is nearly three times the kinetic energy of the fastest α -rays known. Hence we can understand why α particles are never broken up into protons and neutrons as they are hurled about. An alternative possibility, however, is that a helium nucleus might break up into 2 deuterons, or nuclei of heavy hydrogen, H^2 . The mass spectrograph gives

¹ Some writers speak of a conversion of rest mass into energy. The view adopted in this book is that mass accompanies energy of all sorts in an unvarying proportion, so that both mass and energy are always conserved.

$$H^2 = 2.0147 \text{ a.m.u.} \quad (306)$$

Thus

$$2H^2 - He^4 = 4.0294 - 4.0039 = 0.0255 \text{ a.m.u.} = 23.8 \text{ Mev.}$$

This is more than twice as large as the largest α -ray energy, so that He^4 is very stable, also, as against disintegration into 2 deuterons.

Similar results emerge from a study of many other nuclei. Nuclear or atomic masses are often discussed in terms of the difference between the isotopic weight M_i and the mass number A . This difference is commonly called the *mass defect*, although *mass excess* would seem to be a more logical term. The terminology may be extended to the neutron, which may be regarded as a nucleus of atomic number

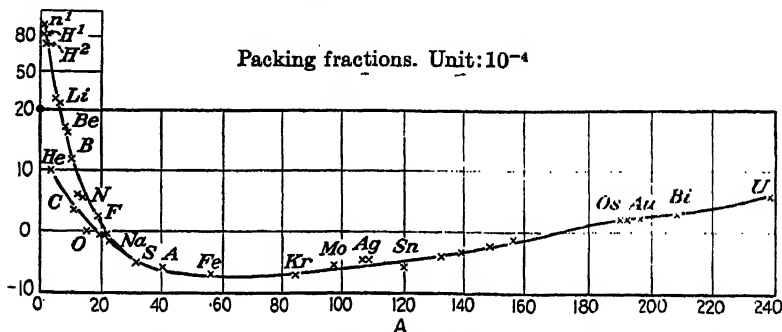


FIG. 198.—Approximate curve showing packing fraction $(M_i - A)/A$ as a function of mass number A (M_i = isotopic weight). The packing fractions of certain isotopes are indicated by crosses.

$Z = 0$. For neutrons, $M_i = 1.0089$, as in Eq. (300), and $A = 1$, so that their mass defect is 0.0089. The mass defect of O^{16} is, of course, zero, because of the way in which mass numbers are defined. The change in mass (*i.e.*, rest mass) in any nuclear reaction is the same as the change in the sum of the mass defects, since the sum of the mass numbers remains constant.

It is more usual, however, following Aston, to discuss the *packing fraction*, or ratio of the mass defect to the atomic number. An approximate curve of packing fractions as a function of mass number A is shown in Fig. 198; the data used in drawing the curve are given in Appendix I. The packing fractions for a few isotopes are indicated by crosses. Below $A = 30$, two curves are drawn, since there is an evident tendency for the packing fraction to be larger when the atomic number is odd than when it is even.

We note that the packing fraction decreases up to about $Z = 60$, after which it increases again. This means that the energy released

per particle in building up the nucleus out of neutrons and protons increases up to $Z = 60$, after which it slowly decreases. As a rough rule, however, we can say that the energy thus released per particle is not very different for atoms beyond boron ($Z = 5$), amounting to something of the order of 0.009 a.m.u., or 8 Mev.

215. The General Theory of Nuclear Binding.—It is clear that the protons and neutrons composing a nucleus must be held together by attractive forces of some sort, otherwise the mutual electrostatic repulsion of the protons would disrupt the nucleus. The true nature of these attractive forces, however, and the exact law of their action are not known with certainty. In this respect the situation in nuclear theory is quite different from that in ordinary atomic theory. The mathematical form of the interaction between electric charges, for some reason, is the same at atomic distances as it is at physically observable distances, the only difference being that at large distances it is possible to use the approximate, classical form of mechanics, whereas on the atomic scale the more exact wave-mechanical form must be employed. Thus, observations of electrostatic interaction on the ordinary physical scale led at once to a knowledge of this interaction on the atomic scale. In constructing a nuclear theory, on the other hand, our only resource is to endeavor to infer the laws of the interaction between the nuclear particles from the observed properties of nuclei, or to try various hypotheses concerning these laws and to compare deductions from them with the results of experiment. We shall not enter upon a detailed discussion of the hypotheses that have been proposed, but shall restrict the discussion to certain ideas that have proved useful and seem destined to form a permanent part of nuclear theory, referring the student elsewhere for further information.¹

One general property that the nuclear forces must possess is indicated by certain features of the scattering of α particles by atomic nuclei. The scattering is found to obey the law deduced by Rutherford (Sec. 100) so long as the distance of closest approach of the α particle to the nucleus, as calculated from its velocity and the angle of scattering, remains greater than a certain distance ρ , which is characteristic of the scattering nucleus. Otherwise departures from Rutherford's formula occur; the scattering is then said to be "anomalous." Since Rutherford's deduction rested on the assumption of the ordinary inverse-square law of force, and since, according to classical theory, this assumption would be justified if the nuclear

¹ Cf. RASSETTI, *op. cit.*; PIERLS, The Theory of Nuclear Forces, *Nature*, vol. 145, p. 687 (1940); BETHE, *Rev. Modern Physics*, vol. 9, p. 69 (1937); and many articles in the literature.

charge were symmetrically distributed within a distance ρ of the center of the nucleus, the value of ρ deduced from the scattering experiments is regarded for convenience as the "radius" of the nucleus. Classical theory is not really applicable to the scattering phenomenon, to be sure, but wave mechanics, also, leads to Rutherford's formula, and it is thought that the classical interpretation of ρ may represent correctly certain features of nuclear theory. The values of ρ deduced from the experiments are represented roughly by the equation

$$\rho = 2 \times 10^{-13} Z^{1/2} \text{ cm.} \quad (307)$$

As indicated by this equation, $\rho \propto Z^{1/2}$, nearly. Thus the *volume of the nucleus*, $\frac{4}{3}\pi\rho^3$, is *proportional to Z*. Since Z , in turn, is nearly proportional to the number of particles in the nucleus, we reach the

conclusion that the *density of all nuclei is nearly the same*. In this respect, there is a strong contrast

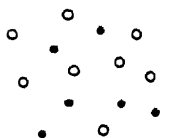


FIG. 199.—
Illustrating a nucleus composed of protons (●) and neutrons (○).

between the nucleus and the electronic envelope of the atom, in which the mean density of the electrons increases rapidly with increasing Z . The fact that the density does not change much when one more particle is added implies that the interactions between particles which hold the nucleus together are *sharply limited in range*, so that the addition of a particle affects only those near to it. The interaction between two particles

must, therefore, drop off quickly to zero when the distance between them exceeds a certain limit; and the observed values of ρ imply that this limiting distance must be of the order of 10^{-13} cm. or less.

Thus a nucleus resembles somewhat a drop of liquid, held together by cohesion. It is doubtless more penetrable, however; perhaps a swarm of bees in the air would be a better analogy. The latter comparison illustrates also the fact that, like the electrons in the atom, the individual neutrons and protons do not have any fixed places in the nucleus. The wave function for the nucleus probably does *not* represent them as *changing position from time to time*, however; it merely assigns a certain probability for each possible position in the nucleus, this probability being the same for each neutron, or for each proton. In Fig. 199 is shown a diagram that may help in visualizing some of these conclusions.

Besides its nearly constant density, a nucleus also resembles a drop of liquid in the fact, noted in the last section, that the binding energy is almost proportional to the number of constituent particles. Just as the heat that is given off when a molecule of vapor condenses onto a

drop of liquid is independent of the mass of the drop, so the energy released when one more neutron or proton is added to a nucleus is (nearly) independent of Z , amounting roughly to 8 Mev per particle.

The classical analogies must not be taken too literally, however. For classical mechanics to be applicable to nuclear phenomena, the energies of the particles would have to be high enough to make their de Broglie wave lengths small as compared with nuclear dimensions. According to Eqs. (166a,b) in Sec. 116, to have a wave length even as small as 10^{-13} cm., a neutron or proton would have to have kinetic energy equal to

$$\frac{1}{2}mv^2 = \frac{(mv)^2}{2m} = \frac{(h/\lambda)^2}{2m} = \frac{(6.61 \times 10^{-27}/10^{-13})^2}{2 \times 1.66 \times 10^{-24}}$$

or

$$1.3 \times 10^{-3} \text{ erg} = 820 \text{ Mev.}$$

This is large as compared with the amount of energy associated with one particle in a nucleus. Presumably, therefore, the correct theory of the nucleus must be wave-mechanical in nature. A nucleus will doubtless possess a characteristic series of quantum states; and, because of the electrical charges on the protons, radiative transitions between these states may occur, accompanied by the emission or absorption of γ -rays. Thus are to be explained the energy levels associated with γ -ray lines, such as those described in Sec. 211.

216. Nuclear Forces and the Ratio of Mass Number to Atomic Number. (a) *The Attraction between Neutrons and Protons and the Neutron-proton Ratio.*—The existence of the deuteron, consisting of 1 neutron and 1 proton, shows conclusively that an attraction of some sort exists between a neutron and a proton. This attraction might conceivably be the sole action holding the nucleus together against the disruptive tendency of the electrostatic repulsion of the protons. It is possible to account for the broad characteristics of stable nuclei on the basis of these two force actions alone [but see (b) below].

Let us neglect at first the repulsion of the protons, supposing only the neutron-proton attraction to exist. Then it is easily seen that all stable nuclei should contain as nearly as possible equal numbers of neutrons and protons, since the energy of a given number A of particles will be a minimum when they are half protons and half neutrons. For, suppose a nucleus contains N neutrons and P protons, and suppose that $N > P$. Then, if a neutron were to be replaced by a proton, the energy of the nucleus would be decreased by the amount of the negative potential energy due to the added proton interacting with $N - 1$ neutrons and increased by the amount of the energy due to the inter-

action of a neutron with P protons. If $N - 1 > P$, the net effect would be a decrease in the nuclear energy. Hence, under these circumstances, there would be a tendency for a neutron to emit an electron as a β particle and to turn into a proton, thereby decreasing the nuclear energy and at the same time making N more nearly equal to P . In the same way, we may infer that, if $P - N > 1$, a proton should emit a positron and turn into a neutron.

The decrease in energy due to such a change would have to be sufficiently great, to be sure, to provide not only some kinetic energy for the emitted particle but also its rest energy mc^2 ; otherwise no change could occur. The energy of interaction of proton and neutron must be relatively large, however. The mass of the deuterium atom H^2 being 2.01472 a.m.u., the loss of mass when it is formed out of a hydrogen atom and a neutron is, by (298) and (300),

$$1.00812 + 1.00893 - 2.01472 = 0.00233 \text{ a.m.u.} = 2.17 \text{ Mev.} \quad (308)$$

In helium the average binding energy of a neutron or proton is considerably greater than this, *viz.*, by Eq. (305),

$$\frac{28.0}{4} = 7.0 \text{ Mev.} \quad (309)$$

For the average loss of energy when a proton or neutron is added to heavier nuclei, the curve of packing fractions indicates something of the order of 8 Mev, as stated in Sec. 214(b). In comparison with such energies, the rest energy of an electron or positron, 0.51 Mev as in Eq. (304), is rather small. Thus, we should expect a strong tendency for N and P to become nearly equal as a result of β -ray activity.

The effect of the *electrostatic repulsion of the protons* must now be considered, however. The potential energy due to this repulsion adds a positive term in the energy, roughly proportional to P^2 . Hence, its presence will cause P to tend to diminish, and the actual ratio of N to P will be that at which there is a balance between the tendency of the proton-neutron attraction to bring N and P into equality and the tendency of the proton repulsion to diminish P .

The magnitude of the energy due to proton repulsion can be estimated from the classical picture of the nucleus described above. Equation (307) indicates that the distance between protons in a nucleus should be of the order of 10^{-13} cm. The electrostatic energy of 2 protons 10^{-13} cm. apart is only

$$\frac{(4.80 \times 10^{-10})^2}{10^{-13}} = 2.30 \times 10^{-6} \text{ erg} = 1.4 \text{ Mev.}$$

This is small compared with the total binding energy of a helium

nucleus, or 28 Mev. From this example it is clear that proton repulsion will have little effect upon the ratio N/P in light nuclei. The repulsive effect will rise rapidly, however, as Z increases. If the 80 protons in a mercury nucleus, for example, were distributed uniformly over the surface of a sphere of 10^{-12} cm. radius, which is a little larger than the value of ρ as given by Eq. (307), their electrostatic energy, computed classically, would be

$$\frac{(Ze)^2}{2r} = \frac{(80 \times 4.80 \times 10^{-10})^2}{2 \times 10^{-12} \times 1.60 \times 10^{-12} \times 10^6} = 460 \text{ Mev.}$$

If they were scattered throughout the sphere, as they are in reality, their energy would be still larger. The total energy released in binding 120 neutrons and 80 protons into Hg^{200} , with a mass of 200.028, is only $120 \times 0.00893 + 80 \times 0.00812 - 0.028 = 1.69$ a.m.u. or 1,570 Mev. Thus, in the heavier nuclei, the effect of proton repulsion becomes comparable with that of the neutron-proton attraction, and the number of protons should, in consequence, be considerably smaller than the number of neutrons in the nucleus.

The composition of the nuclei that occur in nature is in good agreement with these deductions from our tentative assumptions. A glance at the table in Appendix I shows that there is a strong tendency for N , which equals $A - Z$, to equal P or Z up through calcium, $Z = 20$. Beyond this point N always exceeds P , the ratio N/P increasing to about 1.5 in the heaviest nuclei.

The occurrence of *isobars*, or isotopes having the same mass number but different atomic number, might seem to present a difficulty for the theory just stated, since, for a given number A of particles, there should be a unique value of P or Z that makes the energy a minimum. Isobars are not abundant, however, and, where they do occur, usually only two, rarely three, different values of Z occur for a given value of A . Perhaps in such cases the difference in energy is too small to permit transformation, by the emission of a β -ray, into the isobar of lowest energy. The known masses are not sufficiently accurate as yet to permit a general test of this explanation. Furthermore, in a few cases β -ray activity is actually observed to be going on; for example, a rare isotope of potassium, $_{19}\text{K}^{40}$, isobaric with the commonest isotope of calcium, $_{20}\text{Ca}^{40}$, is responsible for the slight radioactivity which potassium has long been known to exhibit.

All known nuclei with $Z > 83$ are likewise radioactive, presumably because of the overwhelming effect of proton repulsion. Probably the nonoccurrence of nuclei with $Z > 92$ is due to the rapidity with which such nuclei, if they existed, would undergo radioactive change.

(b) *The Interactions of Neutrons with Neutrons and Protons with Protons.*—Since a proton and a neutron interact, in spite of the absence of an electric charge on the neutron, it seems probable that neutrons also interact with one other. If this is so, then it may be surmised that a similar short-range attractive interaction may exist also between protons, superposed upon their electrostatic repulsion; it may be that 2 protons, if they come close enough together, will attract each other.

The interaction between protons can be studied by observing the scattering at wide angles of high-speed protons by the element hydrogen, which is just a collection of almost stationary protons so far as the impinging proton is concerned, the deflecting effect of an electron being negligible because of its very light mass.¹ Careful experiments point to the existence of an attractive interaction between 2 protons which can be said very roughly to overbalance their repulsion when they are less than e^2/mc^2 or 3×10^{-13} cm. apart.² (All such statements are essentially of the nature of classical *analogies*, of course; the mathematical analysis of the data follows wave-mechanical lines throughout.)

The attractive effect between 2 protons, as inferred from the scattering experiments just mentioned, is about the same as that between a proton and a neutron. It is natural to suppose that a similar attraction exists also between neutrons. *If the neutron-neutron and proton-proton attractions are equal*, it is easily seen that the conclusions drawn above concerning nuclear composition remain valid. Since these conclusions agree with the facts, we may infer that the two attractive effects, if not equal, are at least not very different.

Another point to consider, in dealing with the binding of particles into a nucleus, is the possibility of *proton and neutron spin*. In the deuteron H^2 , the band spectra of deuterium, H_2^2 , indicate the existence of an angular momentum corresponding to a quantum number $I = 1$. This is easily accounted for if the neutron and proton each has spin with $s = \frac{1}{2}$. In spectroscopic terminology, we may then explain the value $I = 1$ by supposing that the nucleus H_2^2 is normally in a triplet state (cf. Sec. 150).

If these conclusions as to spin are correct, there may well be a tendency for neutrons, and likewise protons, to occur in pairs in the

¹ According to classical mechanics an electron would be set in motion by a proton passing close to it, but, by conservation of momentum, the effect on the velocity of the proton would be very small. Corresponding results follow from wave mechanics.

² Cf. BREIT, THAXTON, and EVERBLUD, *Phys. Rev.*, vol. 55, p. 1018 (1939).

nucleus, the members of each pair having opposite spins. This is in agreement with the observed facts. The smallest nucleus that can be constructed out of completed pairs is the helium nucleus, He^4 . As we have seen, the binding energy per particle of He^4 (7.0 Mev) is of the same order of magnitude as that of all heavier nuclei, and is over three times greater than that of the deuteron (2.2 Mev), which contains only a single neutron and proton. The next three elements after He have smaller binding energies per particle than has He^4 , as is evident from the fact that their packing fractions, given in Appendix I, are larger. Thus the structure of the helium nucleus or α particle must represent a peculiarly stable configuration, in analogy with the closed subshell of atomic theory.

As Z becomes large, the binding energy per particle increases again and finally becomes greater than that of He^4 . This suggests that heavy nuclei may be thought of as formed, in part at least, of helium nuclei, which are more lightly bound together than the particles in the helium nucleus itself. Mass numbers divisible by 4 do not seem to occupy any privileged position, however. On the other hand, there is an obvious tendency for the neutrons and protons to occur in pairs. Inspection of the table in Appendix I shows that elements with even atomic number Z have, as a rule, many more isotopes than those with odd Z . Furthermore, even Z usually goes with an even value of A , the mass number, and odd Z with odd A , and this means that usually the number of neutrons is even. An odd value of Z is combined with an even value of A in only four stable nuclei (H^2 , Li^6 , B^{10} , N^{14}). It is also significant, doubtless, that some 87 percent of the earth's crust is composed of elements of even atomic number.

217. The Explanation of Natural Radioactivity.—The broad features of natural radioactivity are easily understood in terms of the theory of nuclear structure just described, as is also the related fact that elements with $Z > 92$ do not occur in nature. A nucleus containing 93 protons, even if it contained also the most favorable number of neutrons, would presumably break up quickly owing to the repulsions of the protons. In uranium ($Z = 92$) and thorium ($Z = 90$), we see such a process of disintegration actually occurring, although at a slow rate; protactinium (Pa , $Z = 91$) is much scarcer and disintegrates more rapidly, as befits its odd value of Z . Stability is not reached until Z has sunk to 83 (bismuth).

That it should be an α particle that breaks off from a nucleus, under such conditions, rather than either single protons or neutrons or some other group of them, is not so easily seen. This fact must be a consequence of the specially low energy of the two-pair structure of the α

particle. The eventual emission of β particles, on the other hand, is readily explained. As successive α particles are emitted, thereby decreasing equally the number N of neutrons and the number P of protons in the nucleus, the ratio N/P , already greater than unity, becomes still larger; it may become so large that a neutron turns into a proton with the emission of an electron, thereby lowering the energy and bringing N/P nearer to its normal stable value. Thus is explained the occurrence among the naturally radioactive heavy elements of the emission of negative β -rays but never of positive ones, i.e., of positrons.

It is sometimes helpful to employ the following semiclassical picture of the process of α -ray emission. Imagine a group of 2 neu-

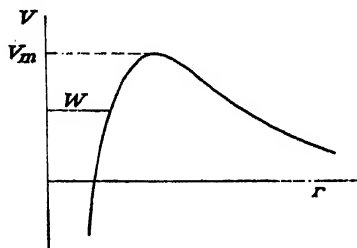


FIG. 200.—Illustrating the “potential barrier” for an α particle in a nucleus.

trons and 2 protons forming an α particle to start out of the nucleus. At first, they will be held back by the attraction of the remainder of the nucleus; as they move out, therefore, the potential energy of the nucleus rises. The attraction rapidly decreases, however, until at a certain point it just equals the repulsion between the α particle and the protons in the remainder of the nucleus; at this

point, the potential energy is a maximum. Beyond this point the repulsion predominates, the potential energy decreases, and the α particle is accelerated outward. The trend of the potential energy V of the nucleus, assumed zero at infinity, as a function of the distance r of the α particle from the center of the nucleus, will thus be as illustrated in Fig. 200. We can think of V as constituting a “potential barrier” tending to hold the particle in the nucleus.

If we adhere strictly to classical theory, to be sure, we encounter serious difficulties in endeavoring to understand the *slow rate* of α -ray emission on the basis of this picture. If an α particle is to escape across the barrier, it must have an energy W exceeding its maximum potential energy V_m when at the top of the barrier. While it is still inside the nucleus, therefore, having presumably the same total energy W but less potential energy, it must possess a considerable amount of kinetic energy and so must be in motion; why, then does it not issue immediately from the nucleus, instead of waiting for an appreciable length of time? Furthermore, why may one atom of a substance like radium disintegrate almost immediately, whereas another atom, apparently identical with the first, may wait a thousand years?

As in so many other cases, wave mechanics comes to the rescue. If $W > V_m$, wave mechanics indicates that the α particle should depart quickly, just as in classical theory. Escape is also possible, however, if $W < V_m$, contrary to the classical result. The process is similar to the Auger effect in atoms [Sec. 201(b)]. Provided $W > 0$, the Ψ waves leak through the barrier, at least at a slow rate, so that the probability of finding the particle outside the nucleus continually increases with the time. The higher the top of the barrier, *i.e.*, the greater the value of $V_m - W$, and the wider the barrier, the slower is the rate at which the wave packet streams outward and the smaller is the chance per second that the α particle issues from the nucleus. In this way, we can understand even such slow disintegration as that of uranium, where the chance of escape of an α particle is only 4.9×10^{-18} per second.

When it does come out, the kinetic energy of the α particle will be equal to its whole energy W . Since, as we have seen, $V_m > W$, the observed kinetic energy of the α particle sets a lower limit to the height of the barrier that must be assumed to surround the nucleus.

ARTIFICIAL TRANSMUTATION OF NUCLEI

Since disintegration of nuclei occurs spontaneously in the case of the radioactive elements, it is natural to wonder whether nuclei might not also be disintegrated by bombarding them artificially with high-speed particles. We have seen, however, that the binding energies of nuclei are at least of the order of several million electron-volts. Until recently the only sources of energies of this order of magnitude were the α particles from the radioactive elements. Recently, other sources of high-speed particles have been developed, and today the transmutation of nuclei is a commonplace in physical laboratories.

218. Artificial Transmutation by Alpha Particles.—It was first shown by Rutherford¹ in 1919 that the nuclei of nitrogen atoms would emit swiftly moving protons when bombarded by the α particles from radium C'. A brief description of Rutherford's experiment will be given, to serve as a basis for a better understanding of what follows. A sketch of his apparatus is shown in Fig. 201. One end of a box B was closed by a metal plate E , in the center of which was a hole covered by a thin sheet of silver foil W . Just outside the window was a zinc sulfide screen, the scintillations on which were observed by a microscope M . The source of α particles, consisting of radium C', was placed on a small disk D , whose distance x from W could be

¹ RUTHERFORD, *Phil. Mag.*, vol. 37, p. 581 (1919).

varied at will. *AA* were side tubes through which various gases could be introduced into the enclosure.

When the box was filled with oxygen or CO_2 at atmospheric pressure, no scintillations were observed on the screen when x exceeded the "range" of the α particles emitted by radium C—in these gases about 7 cm. When, however, the box was filled with nitrogen, scintillations were observed *when x was as great as 40 cm.* Since the α particles from the radium source could not penetrate 40 cm. of air, Rutherford proposed the hypothesis that the α particles, by colliding with the nitrogen nuclei, caused the disintegration of the latter, with the emission of long-range particles which produced the

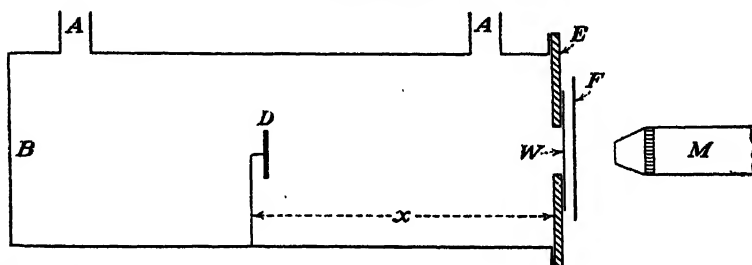


FIG. 201.—Apparatus by which Rutherford observed production of long-range particles when α particles collide with atoms of some of the lighter elements.

scintillations. Measurements of the magnetic deflection of these particles suggested that they were protons—a surmise confirmed by subsequent work. Rutherford and Chadwick¹ later showed that many other light elements could be similarly transmuted. To study the phenomenon in solids, the source *D* was moved close to the window *W*, which was replaced by a thin layer (foil where possible) of the element under study, and the scintillations produced by the ejected protons were observed on the screen *F*. By putting various thicknesses of material of known stopping power between *W* and *F*, the equivalent range in air of the protons emitted from these substances could be determined. In this way it was found that all elements from B(5) to K(19), with the exception of C(6) and O(8), could be transmuted, with the emission of protons.

The ranges of the emitted protons varied from element to element under bombardment by α particles from radium C, as follows, ranges being given in air and expressed in cm.:

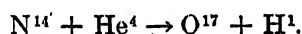
Source.....	B(5)	N(7)	Fl(9)	Na(11)	Mg(12)	Al(13)	P(15)
Range.....	58	40	65	58	40	90	65

¹ RUTHERFORD and CHADWICK, *Phil. Mag.*, vol. 42, p. 809 (1921) and vol. 44, p. 417 (1922); *Phys. Soc., Proc.*, vol. 36, p. 417 (1924).

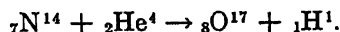
Now Rutherford had measured the range of protons in air as a function of their velocity, and he was able to show that even if *all* of the energy of the α particle from radium C were given to the proton as a result of the collision, it would have a range of only about 57 cm. It is seen, therefore, that in the case of Fl(9), Al(13), and P(15), at least, the initial energy of the proton *exceeds that of the incident α particle*. The disruption of these nuclei must, therefore, be accompanied by an evolution of energy. Whence comes this energy?

Rutherford's answer was that the energy was supplied by the nucleus from which the proton came, this nucleus being caused to "explode" by the impact of the α particle. He assumed that the α particle continued on its way, although with diminished energy and in a different direction. Recent work, on the contrary, indicates that in such phenomena the α particle is actually captured by the nucleus, forming a nucleus of a different element. In a cloud chamber in which the track of the incident α particle, that of the proton, and that of the recoiling nucleus can all be plainly seen, there is no trace of a track made by an escaping α particle.

Rutherford's original nitrogen transmutation is, therefore, now believed to be represented by the following equation:



Here, for simplicity, we have written the symbols usually employed for neutral isotopic atoms to represent their nuclei, He^4 indicating the incident α particle and H^1 the ejected proton. According to the equation, the nitrogen nucleus gains 4 units of mass from the He^4 and loses 1 unit in H^1 , so that its mass rises by 3 units, from 14 to 17; and it also gains 2 units of positive charge from He^4 and loses 1 unit in H^1 , so that Z rises by 1, from 7 to 8. The nitrogen nucleus thus becomes transmuted into a nucleus of the rare isotope of oxygen, O^{17} . In any such equation, the sum of the atomic numbers Z must balance on both sides of the equation, representing conservation of electric charge, and the sum of the atomic numbers A must also balance, representing conservation of neutrons and protons. The balance for Z could be exhibited numerically in the equation by writing it in terms of more explicit isotopic symbols, thus:



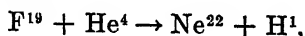
Actually, theoretical considerations indicate that, in such occurrences, a single intermediate nucleus is probably first formed, in a very unstable condition, so that a more complete description of the reaction

would be



The intermediate nucleus cannot (ordinarily) be detected experimentally, however, so it is simplest to omit it in writing the reaction.

There should also be a balance in terms of *energy*. In equating energies it will obviously be necessary to include the rest energy of the nuclei as well as their kinetic energy. In the particular case under discussion, the mass of O^{17} has not been measured directly; therefore, this particular nuclear reaction can only be used the other way round, as a means of calculating the mass of O^{17} . In the case of another of Rutherford's reactions,



all four masses have been measured directly and the energy principle can be used to furnish a test of the assumed reaction. Actually, however, the energy balance involves some other factors such as a possible emission of γ -rays; therefore, further discussion of it will be postponed [see Sec. 225(*k*)].

219. Artificial Transmutation by High-velocity Protons.—During the last 10 years, methods have been developed for producing, artificially, positive ions of high velocity, thus making it possible to conduct transmutation experiments without the limitations incident to the use of radioactive sources. A later section will be devoted to the discussion of such methods. It is appropriate at this point, however, to describe the pioneer experiments of Cockcroft and Walton, in which nuclear transmutations were first produced by means of artificially accelerated ions.¹

The high potential for use in these experiments was produced by means of a combination of condensers and rectifiers.² The method has analogies with the old scheme of charging a number of condensers in parallel, insulating them, and then connecting them in series so as to add up their potentials; but here all connections are made electrically. Two columns of condensers are used, cross connected with rectifiers in such a way that, when the potential of either column is raised sufficiently relative to the other by means of a potential difference applied at their bases, electrical charge flows across through the rectifiers to the other column. Thus, by using an alternating potential difference, charges are pumped up to the tops of the columns. Potentials as large as 700,000 volts could be produced in this manner,

¹ COCKCROFT and WALTON, *Roy. Soc., Proc.*, vol. 137, p. 229 (1932).

² Cf. COCKCROFT and WALTON, *Roy. Soc., Proc.*, vol. 136, p. 619 (1932); GREINACHER, *Zeits. f. Physik*, vol. 4, p. 195 (1921).

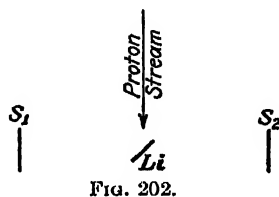
using an applied voltage of a quarter as much supplied by a high-tension transformer.

In this way Cockcroft and Walton were able to bombard a number of the lighter elements with protons having energies up to 500,000 electron-volts, the proton current being of the order of 10^{-6} ampere. Under such bombardment, a number of the elements emitted α particles, the process being thus the converse of bombardment by α particles. When lithium, for example, was bombarded by protons of 250 electron-kilovolts equivalent energy, α particles were observed whose range in air was about 8 cm., corresponding to an energy of some 8.6×10^6 electron-volts. Increasing the bombarding energy up to 500 electron-kilovolts *did not appreciably alter the range of the emitted α particles*. The α particles were observed both by means of scintillations and with an electrical counter [Sec. 209(a),(b)]; later their tracks were observed by Dee and Walton in a cloud chamber.¹

These results were explained by assuming that the proton penetrates the Li^7 nucleus, after which the product nucleus thus formed breaks up into two α particles, the net reaction being



If this interpretation is correct, a simple conclusion can be drawn concerning the directions of projection of the α particles. Their momentum is so large that the momentum



of the incident proton can be neglected; it follows then from the conservation of momentum that the α particles must be projected with approximately *equal velocities in opposite directions*. Cockcroft and Walton confirmed this conclusion by placing a thin lithium target at 45° to the stream of incident protons as shown in Fig. 202. They found that in a large number of cases scintillations appeared *simultaneously* on the screens S_1 and S_2 .

The *energy relations* in this reaction are particularly interesting because the energy of the α particles so greatly exceeds that of the incident protons. The excess energy must be derived from the rest energy of the Li^7 nucleus. As an energy equation we may write

$$\text{Li}^7 + \text{H}^1 + E(\text{H}^1) = 2\text{He}^4 + E(2\text{He}^4),$$

in which isotopic symbols have been written to represent rest energies and $E(\text{H}^1)$ and $E(2\text{He}^4)$ stand for the kinetic energies of the proton and of the α particles, respectively. Or, following recent custom, we

¹ DEE and WALTON, *Roy. Soc., Proc.*, vol. 141, p. 733 (1933).

may write

$$\text{Li}^7 + \text{H}^1 = 2\text{He}^4 + Q \quad (311)$$

where $Q = E(2\text{He}^4) - E(\text{H}^1)$ and represents the *heat of reaction* or net amount of energy released by the reaction, in the form of kinetic energy in the present case. To test the equation, we have, from modern data, using Eq. (303):

	Mev
Rest energy of Li^7 (7.0180 a.m.u.).....	6540.8
Rest energy of H^1 (1.0081 a.m.u.).....	939.6
Kinetic energy of H^1	0.250
Rest energy of He^4 (4.0039 a.m.u.).....	3731.6
Kinetic energy of each He^4	8.6

Thus from the kinetic energies

$$Q = 2 \times 8.6 - 0.25 = 16.95 \text{ Mev};$$

whereas, from the masses and Eq. (311),

$$Q = \text{Li}^7 + \text{H}^1 - 2\text{He}^4 = 17.2 \text{ Mev}.$$

The agreement is within the experimental error. Thus we are dealing here not only with a *transmutation of the elements* but also with a *genuine release of energy out of the lithium atom, at the expense of its ordinary measured mass*.

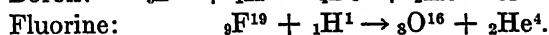
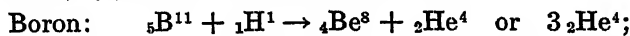
In this calculation we have used masses of the *neutral atoms*, whereas the equation refers, strictly speaking, to the *bare nuclei*. The balance of the atomic numbers in the equations implies, however, a balance in the numbers of circumnuclear electrons as well; hence, errors due to the inclusion of the electronic masses in the values used for nuclear masses cancel out.

To make such calculations significant, it is obvious that atomic masses must be known with high precision; and many of them are. The error in the mass-spectrographic values cited here is supposed to be less than 0.0002 in the case of Li^7 and less than 0.00007 for each of the others. If these errors all had the value stated and the same sign, the error would be 0.0005 or about 0.5 Mev. It is most likely, however, that the error in Q as calculated from the masses would not exceed 0.2 Mev, which is about the difference between the two values of Q as found above.

The kinetic energy of the proton is so small in this case that variation in its energy should make little difference in the range of the α particles, as was actually found to be the case.

Cockcroft and Walton also tested the elements Be, B, C, O, F, Na, Al, K, Ca, Fe, Co, Ni, Cu, Ag, Pb, U and observed an emission of α

particles in most cases. Later work has shown that in many cases the weak effect observed was really due to an impurity in the target. The following nuclear reactions proposed by Cockcroft and Walton, however, appear to have been confirmed:



Thus, in the case of boron, the product is sometimes three α particles, sometimes two α particles and a Be^8 nucleus.

For *fluorine*, the energy equation can be written

$$\text{F}^{19} + \text{H}^1 - \text{O}^{16} - \text{He}^4 = Q,$$

$$19.0045 + 1.0081 - 16 - 4.0039 = 0.0087 \text{ a.m.u.},$$

corresponding to

$$Q = 8.1 \text{ Mev.}$$

Here Q represents the gain in kinetic energy due to the reaction, or the kinetic energy of O^{16} and He^4 minus that of H^1 . According to later, more precise observations,¹ the α particles produced by this reaction have a range of 6.95 cm. when released by protons accelerated in a cyclotron (Sec. 224) to an energy of 1.63 Mev. The kinetic energy given to the O^{16} nucleus was not measured, but allowance for it can easily be made in the following manner.

In the later experiment just mentioned, the α particles were observed in a direction at right angles to the proton beam; and under such circumstances a complete dynamical analysis of the collision is easily made. Let M_1 , M_2 , M_3 denote the masses of the proton, α particle, and residual nucleus, O^{16} , respectively. The O^{16} nucleus moves off obliquely (Fig. 203); let us resolve its velocity into two components, V_{\parallel} parallel to the proton velocity v , and V_{\perp} opposite to the velocity u of the α particle. Then by conservation of momentum

$$M_1 v = M_3 V_{\parallel}, \quad \therefore V_{\parallel} = \frac{M_1 v}{M_3};$$

$$M_2 u = M_3 V_{\perp}, \quad \therefore V_{\perp} = \frac{M_2 u}{M_3}.$$

Thus

$$Q = \frac{1}{2} M_2 u^2 + \frac{1}{2} M_3 (V_{\parallel}^2 + V_{\perp}^2) - \frac{1}{2} M_1 v^2 = \frac{1}{2} M_2 u^2 \left(1 + \frac{M_2}{M_3} \right) - \frac{1}{2} M_1 v^2 \left(1 - \frac{M_1}{M_3} \right)$$

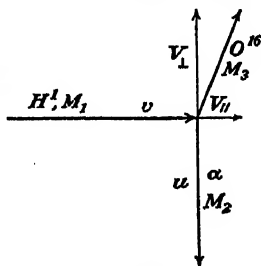


FIG. 203.

¹ HENDERSON, LIVINGSTON, and LAWRENCE, *Phys. Rev.*, vol. 46, p. 38 (1934).

or

$$Q = \left(1 + \frac{M_2}{M_3}\right) E_2 - \left(1 - \frac{M_1}{M_3}\right) E_1, \quad (312)$$

where $E_1 = M_1 v^2/2$ and $E_2 = M_2 u^2/2$, representing the respective kinetic energies of the incident proton and of the ejected α particle.

For 6.95-cm. α particles, we have $E_2 = 7.70$ Mev.¹ Substituting in (312) this value and $E_1 = 1.63$ Mev and the approximate masses $M_1 = 1$, $M_2 = 4$, $M_3 = 16$, we find

$$Q = 8.10 \text{ Mev.}$$

This agrees excellently with the value of Q calculated above from the atomic masses (8.11). Every such check not only confirms the particular nuclear reaction that has been assumed to occur but also

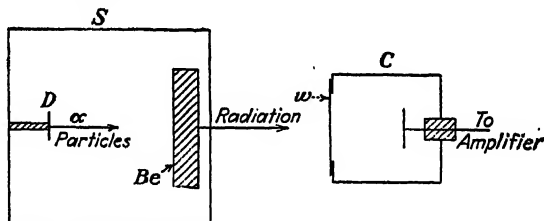


FIG. 204.—Chadwick's apparatus for the production of neutrons.

supports once more the principle of the close relation between mass and energy.

220. Discovery of the Neutron.—Late in 1930, Bothe and Becker observed that a number of the lighter elements, when bombarded by α particles from polonium, emitted a very penetrating radiation. The effect was very marked for beryllium, the radiation from which was reduced in intensity only one half by passing through 2 cm. of lead. An even more penetrating radiation was emitted when the α particles struck boron. It was presumed that these radiations consisted of γ -rays.

In 1932, however, Mme. Curie-Joliot and M. Joliot² observed that these radiations from beryllium and from boron were able to project protons at extraordinarily high velocities from matter containing hydrogen. The experiment was repeated by Chadwick,³ who used the arrangement illustrated in Fig. 204. The polonium emitting

¹ LIVINGSTON and BETHE, *Rev. Modern Physics*, vol. 9, p. 245 (1937). Plot on p. 266.

² CURIE and JOLIOT, *Comptes Rendus*, vol. 194, p. 273 (1932).

³ CHADWICK, *Roy. Soc., Proc.*, vol. 136, p. 692 (1932).

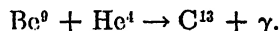
the particles was placed on a disk D in an evacuated vessel S and opposite a disk of beryllium, Be . The radiation emitted by the beryllium entered the window w , of thin aluminum foil, of the so-called "valve counter"¹ C , which registered a count every time a quantum of the radiation entered C . With 3 cm. of air between S and C , some 4 counts per minute were registered. This number was not changed materially by inserting as much as 2 cm. of lead between S and C , which showed the great penetrating power of the radiation.

If, however, a thin sheet of paraffin wax was placed immediately in front of w , the number of counts was markedly increased. This increase Chadwick ascribed to the ejection of particles from the paraffin by the penetrating radiation, some of these particles passing through the window w into the counter and causing it to register. By putting absorbing screens of aluminum just in front of the window, he showed that the particles had a maximum range of some 40 cm. in air; and he concluded that they were protons from the fact that they produced in the counter an electrical impulse of the same size as protons of the same range produced in other ways. A range of 40 cm. for protons means an energy of about 5.7×10^6 electron-volts.

It had been suggested by Mme. Curie-Joliot and M. Joliot that these protons owed their origin to a collision between photons of the penetrating radiation from the beryllium and the hydrogen nuclei in the wax, similar to the Compton effect for electrons. If the hydrogen nucleus recoils, as does the electron in the Compton effect, we can compute, by use of the formula given for E_m in Sec. 198(a), the energy $h\nu_0$ necessary to give the recoiling nucleus 5.7×10^6 electron-volts for the most favorable case where $\theta = 180^\circ$. We find that

$$h\nu_0 = 55 \times 10^6 \text{ electron-volts.}$$

Chadwick found, however, that it was difficult to imagine a nuclear process by which a γ -ray of such high energy could be produced through the impact of an α -ray upon a beryllium nucleus. The most favorable possible reaction would be



The mass of C^{13} was known both from data on the artificial transformation of B^{10} and from observations of the isotope effect in the band spectrum of carbon; the value now accepted is $\text{C}^{13} = 13.00766$. If we subtract this from the modern masses of Be^9 (9.01497) and He^4 (4.00388), we have a net loss of rest mass of .01119 = 10.4 Mev. Adding 5 Mev for the kinetic energy of the α particle and neglecting

¹ CHADWICK, CONSTABLE, and POLLARD, *Roy. Soc., Proc.*, vol. 130, p. 463 (1931).

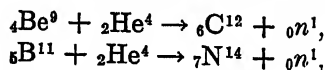
that of C^{12} , we have a maximum of less than

$$15.4 \times 10^6 \text{ electron-volts.}$$

Chadwick's own estimate was 14×10^6 electron-volts, based upon the assumption that for B^9 , whose mass had not then been measured, the mass defect was zero. Both estimates are far below the energy of the supposed γ -ray.

To explain this discrepancy, Chadwick reverted to the "neutron" hypothesis that had been put forward by Rutherford 12 years before. On the older view of nuclei as composed of protons and electrons in close combination, it was reasonable to expect that a nucleus might be formed of a single proton and an electron, combined in much closer fashion than they are in the hydrogen atom. Not being surrounded by an electric field, such a nucleus would produce no ions and so could move with great freedom through matter. It would be deflected or stopped only when it collided bodily with an atomic nucleus, in which case the latter would be projected at a speed proportionate to the speed of the neutron. Many experiments had been made in Rutherford's laboratory in search of such neutrons, but to no avail.

Chadwick assumed that the radiation from beryllium or boron consisted of neutrons projected at high speed according to the reactions



${}_0n^1$ (or n^1 or just n) standing for a neutron. The high penetrating power of the radiation is thus accounted for at once. Furthermore, such particles would require far less energy in order to project a proton at high speed from an atom of hydrogen than would a photon, for a material particle carries much more momentum in proportion to its energy.

According to Rutherford's original hypothesis, the mass of the neutron should be almost the same as that of the proton. This anticipation was confirmed by Chadwick from a comparison of his measurements of recoil protons with certain observations on recoil nuclei of nitrogen produced by the impact of neutrons upon nitrogen molecules. The recoil nucleus thus produced will have a maximum velocity when the collision is central, so that the recoil nucleus is thrown straight ahead and the neutron is thrown straight back. For such a collision, the ordinary laws of elastic collisions give for the velocity v of the recoil nucleus¹

¹ If v'_0 is the speed of the neutron after the collision, $M_0(v_0 + v'_0) = Mv$ by conservation of momentum and $\frac{1}{2}Mv^2 = \frac{1}{2}M_0(v_0^2 - v'^2_0)$ by conservation of energy.

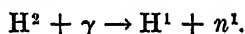
$$v = \frac{2M_0}{M + M_0}v_0, \quad (313)$$

where M and M_0 are the respective masses of the nucleus and the neutron and v_0 is the initial velocity of the neutron. Now, for the protons produced by neutrons, Chadwick deduced from their maximum range a maximum velocity of $v = 3.3 \times 10^9$ cm. sec.⁻¹. For the recoil nitrogen nuclei, Feather, observing the tracks made by these nuclei in a cloud chamber, found a maximum track length of 3.5 mm. in air at 15° and 76 cm. pressure, corresponding, according to previous measurements by Blackett and Lees, to a velocity¹ of $v = 4.7 \times 10^8$. Inserting these values of v in turn in (313), and also $M = 1$ for protons and $M = 14$ for nitrogen nuclei, we obtain

$$3.3 \times 10^9 = \frac{2M_0}{1 + M_0}v_0, \quad 4.7 \times 10^8 = \frac{2M_0}{14 + M_0}v_0,$$

from which $M_0 = 1.2$, $v_0 = 3.0 \times 10^9$ cm. sec.⁻¹

A much better value for the mass of the neutron can be obtained from later observations on the disintegration of the deuteron (nucleus of heavy hydrogen) by γ -rays, which was discovered by Chadwick and Goldhaber in 1934.² In this reaction, the γ -ray photon is absorbed by the deuteron, and the proton and neutron composing the latter fly apart at considerable velocity:



From the size of the pulse produced by the proton in the ionization chamber, Chadwick and Goldhaber estimated the energy given to it in the reaction at 0.24 Mev. Since the velocity of the neutron must be about equal and opposite to that of the proton, in order to preserve the conservation of momentum, the momentum of the photon being much smaller, the total kinetic energy of the two particles is thus close to 0.48 Mev. The γ -rays used, from thorium C'', were very hard, with $h\nu = 2.62$ Mev. The difference,

$$2.62 - 0.48 = 2.14 \text{ Mev} = 0.00230 \text{ a.m.u.},$$

represents energy which must appear in an excess of the combined rest masses of proton and neutron over the rest mass of the deuteron; it represents, therefore, the "binding energy" of the deuteron. From mass-spectrographic data, the deuteron exceeds the proton by

$$2.01472 - 1.00812 = 1.00660 \text{ a.m.u.}$$

¹ FEATHER, *Roy. Soc., Proc.*, vol. 136, p. 709 (1932).

² CHADWICK and GOLDHABER, *Nature*, vol. 134, p. 237 (1934); *Roy. Soc., Proc.*, vol. 151, p. 479 (1935).

Adding 0.00230 to this, we have, finally, for the mass of the neutron

$$M_0 = 1.00890 \text{ a.m.u.}$$

Another reaction discussed by Bethe¹ gives $M_0 = 1.00893$. Thus the neutron is heavier than the proton by $1.00893 - 1.00812 = 0.00081$ or about one and a half times the electronic mass (0.00055).

In subsequent work, neutrons have been found to produce many interesting effects in colliding with nuclei. We shall return to this subject (Sec. 223) after describing *two other fundamental discoveries* that were made about the same time.

221. Positrons.—In the course of some observations of tracks made by cosmic-ray particles in a cloud chamber, in the presence of a magnetic field of some 15,000 oersteds, Anderson, in 1932, obtained a photograph of a track which could only have been caused by a positively charged particle² (reproduced in Fig. 226). The track passed through a sheet of lead that had been placed in the chamber but was unequally curved on the two sides of the lead. This difference of curvature was ascribed, as in other similar cases, to a loss of energy of the particle in the lead, the direction of motion being assumed to be from the side on which the track was less curved toward that on which it was more curved, indicating a lower velocity. If this conclusion as to the direction of motion was correct, however, the *direction* of the curvature of this particular track was such as to indicate that the particle carried a *positive* charge. Yet it could not be a proton; for the curvature was so great that the particle, if a proton, would have an energy of only 300,000 volts, and a proton of this energy is known to have a range in air of only 5 mm., whereas Anderson's track showed no diminution of curvature in a distance of 5 cm. Other tracks of the same nature were observed subsequently. Anderson concluded that these tracks were caused by positively charged particles having the same mass and numerical charge as the electron. He called the new particle, at first, a "positive electron," later a "*positron*." Negative electrons he proposed to call "*negatrons*," thus leaving the term electron to stand for particles of both signs; but "negatron" does not seem to be coming into use, and the meaning attached to the term "electron" varies.

Subsequent work has confirmed Anderson's discovery and has revealed other experimental sources of positrons. They are emitted by many artificial radioactive substances, as described in later sections. Furthermore, they can be produced by the passage of γ -rays,

¹ BETHE, *Phys. Rev.*, vol. 53, p. 313 (1938).

² ANDERSON, *Science*, vol. 76, p. 238 (1932); *Phys. Rev.*, vol. 43, p. 491 (1933).

or even, probably, of high-speed electrons, through nuclear fields. So far as is known, they are an exact counterpart of ordinary electrons, differing only in the sign of their charge. In a cloud chamber, they produce the same thin, beady track, but curved in the opposite direction in the presence of a magnetic field.

Positrons are produced by γ -rays in a manner strikingly different from any effect of radiation so far described. It appears that when a γ -ray of sufficient energy traverses a strong electric field, like that near an atomic nucleus, there is a certain chance that the γ -ray may, in a sense, be converted into an electron and a positron. Thus, algebraic charge is conserved, equal amounts of positive and of negative electrification being produced. This phenomenon is called "pair production." Of the γ -ray energy $h\nu$, a part $2mc^2$ must go into the rest energy of the pair of particles; the excess $h\nu - 2mc^2$ appears as their kinetic energy. Hence, such processes can occur only if the γ -ray photon has energy $h\nu > 2mc^2 = 1.02$ Mev or a wave length

$$\lambda < 0.012 \text{ \AA} = 12 \text{ X.U.}$$

A partially satisfactory theory of this effect can be based on Dirac's relativistic theory of the electron;¹ and such experiments as have been performed appear to confirm the theory.² A photograph obtained by Curie and Joliot, showing a positron-electron pair produced by a γ -ray in the gas of a cloud chamber situated in a magnetic field (a rare event), is reproduced in Fig. 205.³

There are thus three distinct ways in which a photon can cause the emission of an electron from matter: by photoelectric absorption, by the process of Compton scattering, and by pair production. The first and third of these three processes involve absorption of the entire photon; the second process merely weakens it. In the first two pro-



FIG. 205.—A pair consisting of a negative electron and a positron produced by a γ -ray in the gas of a cloud chamber situated in a magnetic field.

¹ Cf. HEITLER, "The Quantum Theory of Radiation," Chap. IV, 1936.

² ANDERSON, *Science*, vol. 77, p. 432 (1933); MEITNER and PHILIPP, *Naturwiss.*, vol. 21, p. 286 (1933); CHADWICK, BLACKETT, and OCCHIALINI, *Roy. Soc., Proc.*, vol. 144, p. 235 (1934); CURIE and JOLIOT, *Comptes Rendus*, vol. 196, p. 1581 (1933).

³ CURIE and JOLIOT, *J. de Physique*, vol. 4, p. 494 (1933).

esses, a preexisting electron is ejected from an atom; in the third, a new electron is created, along with a positron. The relative probability of the three processes depends upon the frequency of the photon and also upon the atomic number of the atom in which the process occurs. A beam of soft X-rays liberates many more photoelectrons than Compton electrons. As the frequency moves into the typical γ -ray region, however, the photoelectric absorption decreases rapidly, and the Compton scattering, which decreases only slowly, becomes dominant. Finally, as $h\nu$ passes $2mc^2$, pair production begins, and ultimately, at extremely high frequencies, it becomes the principal cause of the weakening of a γ -ray beam as it passes through matter. To give a birds-eye view of the facts, there are indicated roughly in Table V, for 4 typical atoms, the photon energies at which successive pairs of the three processes become equally probable.

TABLE V

	C	Al	Cu	Pb
Photoelectric and Compton effect are equal at $h\nu$ (Mev).....	0.025	0.05	0.15	0.5
X.U.....	500	260	85	22
Compton effect and pair production are equal at $h\nu$ (Mev).....	27	15	9	5
X.U.....	0.45	0.8	1.4	2.5

Since positrons can be created, we should anticipate that they can also be destroyed; and such indeed appears to be the fate of all positrons. The commonest process of destruction is one in which a positron unites with a free or loosely bound electron and both disappear, their entire energy passing away in the form of 2 photons of equal frequency. If their kinetic energy is negligible, the photon frequency will be given by $h\nu = mc^2 = 0.51$ Mev, and its wave length will be $0.024 \text{ A} = 24 \text{ X.U.}$ An alternative but much less probable process is the combination of a positron and an electron near a nucleus and the emission of all of their rest energy in a single photon of

$$h\nu = 2mc^2 = 1.02 \text{ Mev.}$$

Most positrons come to rest before being annihilated,¹ and are then unable to penetrate the interiors of atoms; hence, the observed annihilation radiation should consist mostly of the single frequency

¹ Cf. BETHE, *Roy. Soc., Proc.*, vol. 150, p. 129 (1935).

$\nu = mc^2/h$. This latter radiation has been observed by several experimenters.¹

The older observations of Gray and Tarrant² also merit a comment. They noticed that the very hard γ -rays from Th C'' were scattered very strongly by lead and, upon investigation of the scattered rays, found in them a strong component of about $h\nu = 5 \times 10^6$ ev and also some indication of rays near 10^6 ev. These rays they ascribed at first to fluorescence by the nucleus. It is now believed that they arise from annihilation in the lead of positrons which are produced in the lead itself by the primary γ -rays, for which $h\nu = 2.6 \times 10^6$ ev; the mean free path of a positron so produced in lead should be only of the order of 0.5 mm. We have here an interesting example of the interplay that often occurs between different and apparently unconnected lines of research. Had the positron been discovered a few years earlier, Gray and Tarrant would not have wasted time elaborating their theory of nuclear fluorescence. On the other hand, had the discovery been delayed longer, the theory of nuclear fluorescence might have come to be accepted for a time.

That positrons are not commonly met with in the laboratory is due, no doubt, to the abundance of (negative) electrons in matter with which they may unite. But why are electrons so abundant? Perhaps because there are so many nuclei, built out of neutrons and positive protons, which serve to bind the electrons into atoms. Then why are there no *negative* protons which might similarly bind positrons into atoms? Thus does each discovery in physics raise fresh questions, answers to which it is left for future physicists to discover.

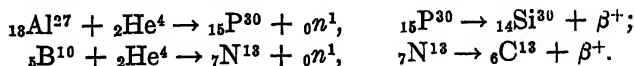
222. The Discovery of Induced Radioactivity.—Almost immediately after Anderson announced his discovery of positrons, it was reported by several experimenters that positrons are emitted from many targets when bombarded by high-speed ions. It was supposed at the time either that the positrons were produced as a direct effect of the bombardment (which is now believed never to occur), or that the bombardment gave rise to γ -rays and these produced pairs of electrons and positrons.

While studying with an ionization chamber the positrons emitted from aluminum or boron when bombarded with α particles from polonium, Curie and Joliot made the important discovery that the emission of positrons did not cease at once when the polonium was

¹ *E.g.*, by KLEMPERER, *Cambridge Phil. Soc., Proc.*, vol. 30, p. 347 (1934); CRANE and LAURITSEN, *Phys. Rev.*, vol. 45, p. 430 (1934).

² GRAY and TARRANT, *Roy. Soc., Proc.*, vol. 136, p. 662 (1932); vol. 143, pp. 681, 706 (1933).

removed.¹ The positron activity was found to decay exponentially, as does the activity of a radioactive substance, with a period (or time required for a decrease to half) of 3 min. 15 sec. in the case of aluminum, or 14 min. in the case of boron. It was concluded that the α particles had transformed atoms of aluminum or boron into atoms of a radioactive isotope of some other element, which then eventually disintegrated into positrons and stable atoms of some sort. The following reactions were suggested, β^+ denoting a positron:



Here Si^{30} and C^{13} are stable isotopes known to occur in nature, but P^{30} and N^{13} are not, only P^{31} , N^{14} , and N^{15} being found. In these cases, the primary reaction results also in the ejection of a neutron.

Now this explanation can at once be put to a chemical test; for P^{30} should be chemically separable from Al, and N^{13} from B. Curie and Joliot made such tests for the aluminum reaction as follows. Aluminum which had been thoroughly activated by α -ray bombardment was dissolved in HCl and the solution evaporated to dryness, within the space of 3 min., and then the residue was tested for positron activity. It showed none. The activity had presumably gone off in the gas that was evolved. It is known that phosphorus treated in this manner would pass off as the gas PH_3 . The process was then repeated, but this time the gas was caught in a tube with walls sufficiently thin to transmit the positrons. Most of the activity was then found in the gas. As a second test, the aluminum was dissolved in aqua regia, without evolution of gas, and a compound of zirconium was added which would precipitate phosphorus as zirconium phosphate, the aluminum remaining in solution; the positron activity was now found in the precipitate but not in the solution. These tests showed conclusively that the substance emitting the positrons was definitely not aluminum and was almost certainly an isotope of ordinary phosphorus. Similar tests identified the substance produced from boron as nitrogen.

The discovery of induced radioactivity advanced nuclear work enormously, because it opened a new possibility of establishing the chemical nature, *i.e.*, the value of the atomic number Z , of the products of nuclear reactions. If these products are radioactive, they can be followed in chemical reactions by means of their radioactivity; whereas otherwise they cannot be located and their chemical nature remains necessarily an inference from indirect evidence, such as considerations

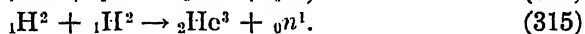
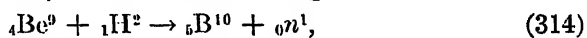
¹ IRENE CURIE and F. JOLIOT, *Comptes Rendus*, vol. 198, pp. 254, 408, 559 (1934).

of mass and energy. It is particularly easy to demonstrate that a given radioactive substance is *not* isotopic with a given element; it is only necessary to mix the two and to show that they precipitate in different proportions, or otherwise behave differently, in some one chemical operation.

As soon as Curie and Joliot announced their discovery, the study of induced radioactivity was taken up in many other laboratories, and new examples of it were announced at a rapid rate. The subject will be discussed in detail in a later section (Sec. 225). Before discussing nuclear reactions further, however, it will be advantageous to return first to a more detailed discussion of the properties of neutrons and then to describe the principal methods that have been developed for the production of high-speed ions to be used in bombarding nuclei. The various lines of nuclear research during the decade from 1930 to 1940 have been so interwoven that they are hard to separate.

223. Experiments with Neutrons.—The discovery of neutrons by Chadwick in 1932 was followed at once by numerous other investigations.¹ Neutrons are much harder to work with than charged particles, since they do not interact with the electrons in atoms and so produce no ions. For this reason their penetrating power is enormous, as we have seen, and they can be detected only by means of nuclear reactions. It has been necessary to invent novel methods for handling them. In this section, a short account will be given of the properties of neutrons as revealed by experiment.

(a) *Sources of Neutrons.*—The original reaction discovered by Chadwick is still widely utilized for the production of neutrons. Radon gas obtained from radium is mixed with finely divided beryllium, from which neutrons are liberated by the α particles from the radon. Such a source lasts only a few days, however; and radon also emits γ -rays, which may be a source of inconvenience. Hence, polonium is sometimes substituted, which emits no γ -rays and has a period of 136 days. Another good method of obtaining neutrons is by bombarding a light element with deuterons from a cyclotron (see next section). The most useful target for this purpose is either beryllium or a substance containing deuterium, such as "heavy" ice or heavy paraffin, the respective reactions being



The $\text{H}^2 + \text{H}^2$ (or $d - d$) reaction requires unusually low voltages; 200,000 volts is enough. When 500,000-volt deuterons fall on heavy

¹ Cf. RASSETTI, *op. cit.*, pp. 233-267; a summarizing article by Amaldi and Fermi, *Phys. Rev.*, vol. 50, p. 899 (1936); and many articles in the literature.

ice, 2×10^7 neutrons per second are produced for each microampere of deuterons absorbed in the target. At higher voltages, however, the beryllium target is superior; from it, a microampere of 1 million-volt deuterons yields 17×10^7 neutrons per second, or 2.7×10^{-5} neutron for each deuteron absorbed. This is about the same yield per particle as in the production of neutrons from beryllium by bombardment with α particles. In any case, as judged by the usual standards, the efficiency of neutron sources is very low.

(b) *The Measurement of Neutron Intensities.*—There are several different methods by which neutrons may be detected and counted.



FIG. 206.—Disintegration of a nitrogen nucleus by a neutron, presumably coming from below. Tracks of the ejected particle and of the recoil nucleus are visible. (*Feather, loc. cit.*)

We may count the recoil protons or other recoil nuclei produced by them in matter; or, if the recoil nuclei are produced in a gas, the ionization caused by them may be measured in an ionization chamber.

Another method is to utilize one of the many nuclear transmutations which neutrons have been found to cause. If such a transmutation occurs in a cloud chamber and results in the ejection of a charged particle from a nucleus, a V-shaped double track will be seen, caused by the charged particle and the residual nucleus, the neutron itself producing, of course, no track (Fig. 206). Or, ionization caused by the products of the reaction may be measured. Ionization chambers lined with lithium or boron, or filled with BF_3 gas, are often used, especially for slow neutrons, the ionization being due to α particles released in one of the two reactions,



(The residual nucleus formed in the first of these reactions is the rare isotope H^3 of hydrogen.)

Finally, Fermi and others have often employed as detectors of neutrons some substance, such as silver, which is made radioactive by them, the induced β -activity of the detector being measured with an ionization chamber. Some of the methods of producing and detecting neutrons are illustrated in Fig. 207.

(c) *Neutron velocities* are most accurately determined by measuring the ranges of recoil protons or other nuclei produced by them, as in the pioneer experiments of Chadwick, described in Sec. 220. The

proton ranges can be measured either with an ionization chamber or in a cloud chamber.

The neutrons projected from a given target bombarded with a homogeneous parallel beam of ions have different velocities according to the direction in which they leave the target. Relative to the composite nucleus that is first formed by the capture of a bombarding ion, the neutrons should all have the same velocity; but this nucleus is given a forward component of motion by the impact. Hence, the neutrons that are projected in the same direction as the bombarding ions have the highest velocity, those projected in the opposite direction the lowest. For experimental purposes it is customary to utilize those neutrons which are projected at right angles to the ion beam, and values of neutron energies are commonly quoted for this direction of emission.

As an example of the results, the neutrons emitted by the " $d-d$ " ($H^2 + H^2$) reaction have been found to form a single, nearly homogeneous group¹; presumably if the bombarding deuterons all had exactly the same energy, so would the neutrons (in a given direction), in accordance with the equation of nuclear energy balance. Neutrons from the Rn - Be reaction, on the other hand, are found to consist of several groups, with a maximum around 4.8 Mev but ranging up to 13 Mev.

(d) *Transmutations Caused by Neutrons.*—Whereas charged particles require high energies in order to penetrate heavy nuclei against electrostatic repulsion, neutrons, carrying no charge, ought to be able to penetrate all nuclei with ease. They ought, therefore, to be especially effective agents for the transmutation of atoms. This expectation is borne out by the facts. Particularly interesting is the *induced radioactivity* produced in many substances by neutron bombardment, first observed by Fermi in 1934.² The induced activity consists of the emission of electrons or, less often, of positrons, accompanied sometimes by γ -rays. The period of decay varies greatly

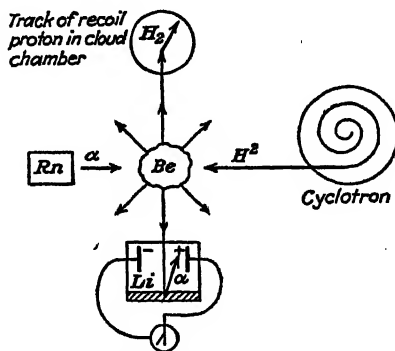


FIG. 207.—Diagrammatic representation of various methods of producing and detecting neutrons.

¹ Cf. HUDSPETH and DUNLAP, *Phys. Rev.*, vol. 57, p. 97 (1940).

² Cf. FERMI, AMALDI, D'AGOSTINO, RASETTI, and SEGRÉ, *Roy. Soc., Proc.*, vol. 146, p. 483 (1934), and references there given.

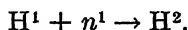
from element to element. Examples of such reactions will be given later (Sec. 225).

(e) *Slow Neutrons*.—The important observation was made accidentally by Amaldi, Fermi, and others¹ that the activity induced by neutrons in silver, rhodium, and many other elements was greatly enhanced when the neutron source and the target were surrounded by material containing hydrogen, such as paraffin or water, or when such material was merely placed between the source and the target. Fermi explained this effect by assuming (1) that in the hydrogenous material the neutrons were slowed down as the result of successive impacts with hydrogen nuclei and (2) that many substances absorb *slow* neutrons much more strongly than they absorb fast ones.

These assumptions have been confirmed by later work. Neutrons have come to be called "slow" when their energies are less than about 1,000 electron-volts. A neutron will lose energy upon colliding with any nucleus that is moving with smaller momentum, but the loss is especially large in collisions with protons because of their relatively small mass. It can be shown that neutrons colliding at random with protons having much smaller velocities will have their energies reduced on the average in the ratio $1/e$ by each collision, e standing for the Napierian base. Thus, a neutron entering hydrogenous material at room temperature with an energy of 5 Mev, provided it does not combine with some nucleus, will be reduced to the energy of a molecule at room temperature, or 0.025 ev, after some 20 collisions (since $\log_e (5 \times 10^6/0.025) = 19.1$); thereafter its energy will oscillate about this latter value. Neutrons which have thus come into thermal equilibrium with matter are called *thermal neutrons*.

The mean free path of neutrons between elastic collisions in paraffin or water has been found to be of the order of 5 cm. at 5 Mev, but it rapidly decreases, until for a thermal neutron it is only 3 mm. Thus, neutrons that have passed through 10 cm. of paraffin will mostly be thermal neutrons. That their velocities really are only of thermal magnitude has been shown by mechanical methods.²

There is, to be sure, a slight chance that a neutron may actually combine with a proton during the slowing-down process, according to the reaction



¹ AMALDI, D'AGOSTINO, FERMI, PONTECORVO, RASETTI, and SEGRÉ, *Roy. Soc., Proc.*, vol. 149, p. 522 (1935).

² DUNNING, PEGRAM, FINK, MITCHELL, and SEGRÉ, *Phys. Rev.*, vol. 48, p. 704 (1935); RASETTI, SEGRÉ, FINK, DUNNING, and PEGRAM, *Phys. Rev.*, vol. 49, p. 104 (1936).

In this reaction, the excess energy is radiated away as γ -rays, which have been detected. The probability of capture increases as the velocity of the neutron decreases, but even at thermal velocities the neutron will execute many elastic collisions with protons before being captured—70 to 120 according to Bethe's theoretical estimate,¹ about 150, occupying a time of 1.6×10^{-4} sec., according to experimental indications.²

Presumably the ultimate fate of all neutrons is to be captured by some nucleus. Slow neutrons are especially interesting because for them the probabilities of capture in going a given distance are relatively large.

(f) *Scattering and Absorption of Neutrons.*—A neutron passing close to a nucleus may be either captured or merely "scattered," *i.e.*, deflected and perhaps changed in energy as if by an elastic collision. The probability of either event is commonly expressed in terms of an equivalent *cross section*. The *scattering cross section* of a nucleus for neutrons can be defined as follows. Imagine a uniformly distributed beam of neutrons, all moving in the same direction and at the same speed, to pass over the nucleus. A certain number of them will be deflected or "scattered" as the result of an elastic collision with the nucleus. This number is the same as the number which, according to classical conceptions, would cross an area of a certain size drawn on a plane perpendicular to the beam. The latter area is the cross section for scattering of neutrons by the nucleus in question. In a similar way, a *cross section for capture* of a neutron by a given nucleus can be defined.

The *scattering cross section* for neutrons is found to increase with the atomic number of the nucleus, but to vary only moderately (in general) with neutron speed. Very roughly, it may be said that about 5 cm. of any solid or liquid material will scatter half of the neutrons out of a beam.

Cross sections for capture, on the other hand, vary greatly, and irregularly, from one element to the next; and they also vary a great deal with the speed of the neutron. Capture cross sections are usually relatively small for fast neutrons and relatively large for thermal ones; for this reason, neutrons to be used for the production of transmutations are commonly first slowed down in paraffin. A few values of the capture cross section σ_c for thermal neutrons at room temperature are, in units of 10^{-24} cm.²:

¹ BETHE and BACHER, *Rev. Modern Physics*, vol. 8, p. 129 (1936).

² Cf. RASSETTI, *op. cit.*, p. 261.

Z.....	3	5	15	47	48	49	64	79	80
Element.....	Li	B	P	Ag	Cd	In	Gd	Au	Hg
σ_c	65	600	1	55	2,900	300	22,200	90	450

Some of these cross sections seem, at first sight, to be astonishingly large in comparison with estimates of nuclear size. Equation (307) gives for the radius of the cadmium nucleus, for example,

$$\rho = 2 \times (48)^{1/3} \times 10^{-13} = 7.3 \times 10^{-13} \text{ cm.}$$

and, for its area of cross section, $\pi\rho^2 = 1.7 \times 10^{-24} \text{ cm.}^2$. Thus for cadmium σ_c is over 1,000 times $\pi\rho^2$. And yet nuclear forces are supposed to have a very short range! The explanation undoubtedly is to be found in the fact that classical mechanics is inapplicable here; we are dealing with a wave phenomenon, not a collision between classical particles. In Sec. 51 it was remarked that, according to classical electromagnetic theory, an oscillating electron, in spite of its minute size, might absorb as much radiation as falls on an area equal to a considerable part of a square wave length. The same principle holds in wave mechanics. The de Broglie wave length of a thermal neutron, by Eq. (166a) in Sec. 116, is $1.5 \times 10^{-8} \text{ cm.}$, so that even $1/100$ of a square wave length, or $2 \times 10^{-18} \text{ cm.}^2$, is far larger than any known value of σ_c .

Wave-mechanical theory indicates, further, that the *cross section for capture* should vary *inversely as the speed* of the neutron, *i.e.*,

$$\sigma_c \propto \frac{1}{v}.$$

It is as if the chance of capture were proportional to the length of time during which the neutron remains in the neighborhood of the nucleus. For velocities of thermal magnitude, this law has been shown by experiment to hold in the case of silver and boron, but it does not hold for cadmium. In these experiments, the velocity of the neutrons relative to the target was varied by causing the target, mounted on a revolving disk, to have a known velocity toward or away from the neutron source.¹

The absorption cross section of cadmium does decrease, however, if the neutron velocity is increased sufficiently. A layer of cadmium 0.5 mm. thick effectively absorbs all thermal and near-thermal neutrons out of a beam; in the process of absorption, the cadmium atoms are increased in mass number by 1, the excess energy being emitted

¹ *Phys. Rev.*, vol. 49, pp. 104, 777 (1936).

as γ -radiation. The absorption is very much weaker, however, for neutrons of energy above 0.4 electron-volt. For this reason, cadmium screens are often used where it is desired to filter out neutrons of thermal energies and to leave only the faster ones. In Fig. 208 is reproduced a curve obtained recently by Baker and Bacher showing the variation of σ_c in cadmium for neutrons of low energy.¹

The absorption by capture of neutrons in cadmium is thus *selective* in much the same way as the absorption of light by dyes is selective. A number of other substances seem to possess absorption lines or bands, commonly called "resonance" levels of energy, mostly at neutron energies of 1 or a few volts. The phenomenon is of great theoretical interest. Wave mechanics predicts the proportionality of

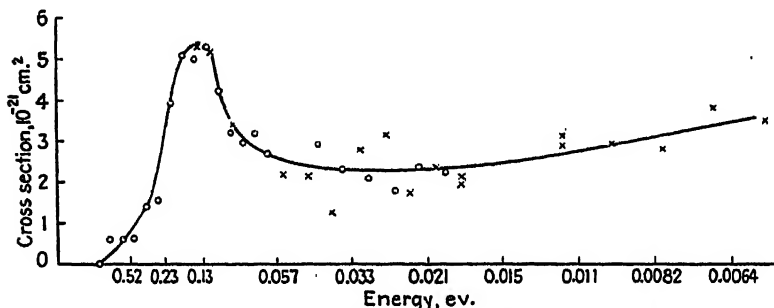


FIG. 208.—Variation of the cross section for capture of very slow neutrons in cadmium. (After Baker and Bacher.)

σ_c and $1/v$ only for neutron energies that are small relative to the lowest resonance level. The experimental facts are not very clear as yet (1940), however, because of the difficulty in working with monochromatic neutrons of adjustable velocity.

Some typical nuclear reactions caused by neutrons will be discussed below. First, however, it will be appropriate to turn aside for a moment from the main narrative and to describe the devices that have been developed during the decade 1930–1940 for producing high-speed ions for use in nuclear research.

224. The Production of High-speed Ions.—In view of the high energies that are involved in natural radioactive processes, taken together with Rutherford's success in transmuting nuclei by bombardment with α particles, it was early realized that the study of nuclear phenomena would be greatly advanced if the experimenter had available a copious supply of ions of various sorts moving at very high and adjustable velocities. From about 1930 to the present, many

¹ BAKER and BACHER, *Phys. Rev.*, vol. 59, p. 332 (1941).

methods of producing such ion beams have been tried.¹ In most of them the ions were accelerated to their final velocities in a long vacuum tube, to the terminal electrodes of which there was applied a potential difference of the necessary magnitude.

Mention has already been made of the condenser-rectifier cascade used by Cockcroft and Walton for the production of large potential differences (Sec. 219). Transformers with their secondary windings connected in series have also been employed for the production of voltages up to 1 million volts; to avoid having a potential difference of this magnitude between the primary and secondary of a single transformer, part of the secondary winding of each transformer is used to supply voltage for the primary of the next, only the first transformer being supplied from an independent source. A disadvantage of this method is that the final voltage produced is necessarily subject to fluctuations whose frequency is that of the alternating current.

A device free from this latter objection is the electrostatic generator developed by Van de Graaff.² The principle is illustrated in Fig. 209. An endless belt B is caused to run rapidly over two pulleys placed one above the other. Electric charge from a generator G is sprayed onto the lower end of the rising half of the belt by means of corona discharge from the points of a comb C . At the top, the belt enters a spherical terminal T , to which the charge is transferred from the belt by means of another comb C' ; or, still better, charge of the opposite sign is sprayed onto the belt, supplied by another generator G' inside the terminal. In this way the charge on the terminal is continually built up, the potential attained being limited only by the rate at which it leaks off. A discharge tube can then be connected between the terminal and the ground, or between the oppositely charged terminals of two such generators standing side by side.

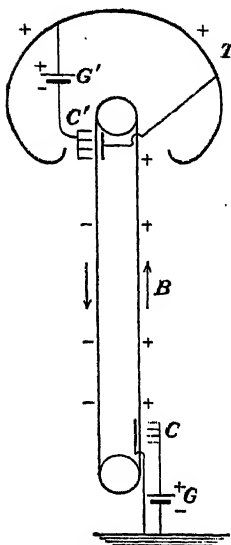


Fig. 209.—Diagram illustrating the principle of the Van de Graaff generator.

¹ For a more complete discussion than can be given here see Livingston and Bethe, *Rev. Modern Physics*, vol. 9, p. 245 (1937), also two articles in *J. of Applied Physics*, vol. 9 (1938), one by Wells on p. 677 and another by Kurie on p. 691.

² VAN DE GRAAFF, *Phys. Rev.*, vol. 38, p. 1919 (1931); vol. 43, p. 149 (1933); L. C. VAN ATTA, NORTHUP, C. M. VAN ATTA, VAN DE GRAAFF, *Phys. Rev.*, vol. 49, p. 761 (1936).

The Van de Graaff generator constructed by the Massachusetts Institute of Technology yields currents of 4 milliamperes at 2.5×10^6 volts.¹ A similar generator at the Westinghouse Research laboratories yields a current of comparable magnitude at voltages up to 3.7×10^6 volts, the voltage holding steady within $\frac{1}{2}$ percent for several minutes at a time.² This latter generator is surrounded by air at 120 lb. in.⁻² pressure, in order to increase the insulation; it is enclosed within a vessel 47 ft. high, made of steel $\frac{5}{8}$ to $1\frac{3}{4}$ in. thick and shaped like a pear standing on its smaller end. Two belts $31\frac{1}{2}$ ft. long and 18 in. wide are used, running at speeds of about 80 ft. per sec. The construction of such an instrument presents many technical problems and well illustrates the elaborate character of the apparatus that is needed for much of modern physical research. To smash a nucleus 10^{-14} in. in diameter, the physicist constructs a machine standing 50 ft. high!

The *vacuum tube* used for the acceleration of ions by means of high potentials is commonly made in sections, and inside it a series of insulated metal cylinders are mounted with short gaps between them (Fig. 210). The ions travel along the axis of these cylinders and the electric field in the gaps exerts a focusing action tending to prevent spreading of the beam away from the axis. In the Westinghouse installation just mentioned, the accelerating tube is about 25 ft. long and consists of 130 sections. The necessity of constructing an accelerating tube long enough to withstand the total potential difference is a drawback to any of the methods so far described for the production of high-speed ions.

A radically different principle is utilized in the well-known *cyclotron* invented by Lawrence and Livingston.³ Instead of attempting to produce a high potential, they accelerate the ions by means of many *successive impulses*. Use is made of the fact that a charged particle moving in a uniform magnetic field revolves in a circle with an angular velocity depending only upon its ratio of charge to mass and upon the strength of the field, but *not upon its velocity*.

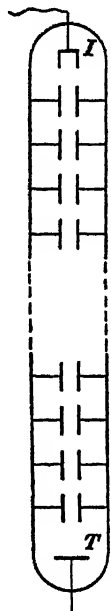


FIG. 210.
Simplified and shortened diagram of an accelerating tube for use with high potentials; *I*, ion source; *T*, target.

¹ *Phys. Rev.*, vol. 57, p. 563 (1940).

² *Phys. Rev.*, vol. 53, p. 162 (1940).

³ LAWRENCE and LIVINGSTON, *Phys. Rev.*, vol. 45, p. 608 (1934); LAWRENCE and COOKSEY, *Phys. Rev.*, vol. 50, p. 1131 (1936). Cf. KURIE, *J. of Applied Physics*, vol. 9, p. 691 (1938).

The accelerating system of a cyclotron is shown diagrammatically in Fig. 211. *BB* represents the horizontal cross section of a flat circular box which can be evacuated. Above it and below it, and only a few inches apart, are the pole pieces of a powerful electromagnet, not shown in the figure; this magnet produces a nearly uniform field of 10,000 to 20,000 oersteds. Inside the box *BB* are two hollow, semicircular electrodes often called "dees," such as might be formed by cutting another circular box in two along the diameter *EE* and slightly separating the halves. Between the dees, a rapidly alternating potential difference, of from 3 to 9 kilovolts at 8 to 15 megacycles, is produced by means of an electrical oscillator.

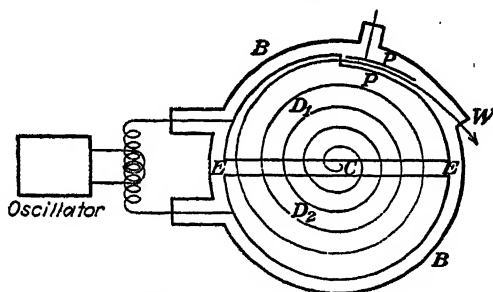


FIG. 211.—Diagram of the essentials of a cyclotron, except for the magnet, whose pole pieces would be situated just above and just below the box *B*.

To produce ions, a suitable gas is introduced into the box at a pressure of 10^{-2} to 10^{-3} mm. Hg, and electrons are shot into this gas at the center of the dees (*C*), from a filament placed just above the dees and maintained at a negative potential relative to the mean potential of the dees. A positive ion thus produced at *C* will be drawn into whichever dee happens to be negative at the moment and will move along a semicircular path in the space inside the dee, where there is no appreciable electric field. If the frequency of the oscillator and the magnetic field strength are properly adjusted, the ion will arrive again at the crack between the dees just at the moment when the potential difference between them has the opposite sign and will be accelerated further. The radii of the successive semicircular paths increase with the speed of the ion, but the time required to traverse a semicircle remains constant and equal to the half-period of the oscillator. Finally the ion, after 50 to 100 revolutions, is traveling near the periphery of the dees and enters the space between two deflecting plates *PP*, where an electric field deflects it slightly outward so that it passes out of the box *BB* through the window *W*. Or, instead of the window, a target may be inserted at this point to be bombarded by the ions.

Naturally a good many accessories are required, and a cyclotron is a fairly large, complicated, and expensive instrument. The dees and pole pieces are commonly 30 in. or more across, so that many tons of iron and copper are required for the water-cooled electromagnet. With such a cyclotron a current to the target of 10^{-5} amp. of deuterons at 5 million volts is readily secured. From the new 60-in., 220-ton cyclotron at the University of California a current of 10^{-5} amp., consisting of deuterons of 16 million volts energy, has been brought out into the air, where it formed a blue streamer nearly 5 ft. long.¹

A limit to the ionic energies that can be produced with a cyclotron will probably be set, ultimately, by the relativistic variation of mass with velocity, which tends to throw the ion out of step with the alternating potential on the dees. A proton with a kinetic energy of 10 Mev has a velocity, calculated nonrelativistically, of

$$\sqrt{\frac{2eV}{m}} = \left(\frac{2 \times 4.8 \times 10^{-10} \times 10^7}{.300 \times 1.66 \times 10^{-24}} \right)^{1/2} = 4.4 \times 10^9 \text{ cm. sec.}^{-1}$$

At this speed the relativistic correction on the mass is already

$$\frac{v^2}{2c^2} = \frac{(4.4/30)^2}{2} = 0.011 \text{ or 1 percent.}$$

For the acceleration of *electrons*, whose speed is close to that of light at energies above 1 Mev, the cyclotron is (at present) quite useless.

225. Some Typical Nuclear Reactions.—Nuclear reactions may be classified either according to the nature of the bombarding particle or ray, or according to the nature of the resulting products. For the bombardment of nuclei, four principal kinds of projectiles have been used: protons, deuterons, helium nuclei (*i.e.*, α particles, or “helions”), and neutrons. Each of these four kinds of particles has been observed on occasion to cause the emission of each of the others, with the single exception that deuterons have been observed to be emitted only in the reaction



The heavier nuclei are more easily transmuted by neutrons, which are not repelled by the nuclear charge; α particles, on the other hand, are effective chiefly on the lighter nuclei. Furthermore, incident γ -rays are observed to cause the emission of neutrons from many materials; and γ -rays are often observed to be emitted as the result of bombardment with material particles.

¹ LAWRENCE *et al.*, *Phys. Rev.*, vol. 56, p. 124 (1939); photograph in an article by Mann, *Nature*, vol. 143, p. 583 (1940).

A bombardment process is often denoted by parentheses enclosing two symbols, the first symbol denoting the incident particle or ray, the second, the emitted particle or ray (or particles); α , p , d , n , γ are used to denote, respectively, an α particle, a proton, deuteron, neutron, or γ -ray. The incident particle is always understood to be captured by the initial or bombarded nucleus. A symbol denoting the initial nucleus may precede the parentheses, and another denoting the residual nucleus left after the departure of an emitted particle (or particles) may follow it. The reaction mentioned in the last paragraph would thus be denoted by $\text{Be}^9(p, d)\text{Be}^8$. In this notation the best known types of nuclear reactions may be listed as follows:

(α, p)	(p, α)	(d, α)	(n, α)	(γ, n)
(α, n)	(p, n)	(d, p)	(n, p)	
	(p, γ)	(d, n)	$(n, 2n)$	
		$(d, 2n)$	(n, γ)	

Here $2n$ denotes the emission of 2 neutrons.

Reactions of all of these types have been found, in certain instances, to result in the production of *radioactive residual nuclei*. A list published in 1940¹ contains 335 different kinds of artificial radioactive atoms, differing from each other in mass number or in atomic number or both. Such atoms are particularly easy to identify because they can be followed by means of their radioactivity and identified by chemical procedures. A radioactive isotope, natural or artificial, is now known for every element except hydrogen and perhaps $Z = 85$. The induced activity consists of the emission of a β particle, either positive or negative; sometimes, however, it is a *secondary* negative β particle, ejected from the K shell of the atom by interaction with an excited nucleus. In some cases the activity is accompanied by an emission of γ -rays. Observed periods (time for the activity to decrease by one half) range from 0.022 sec. (B^{12}) to 7 years (Co^{60}).

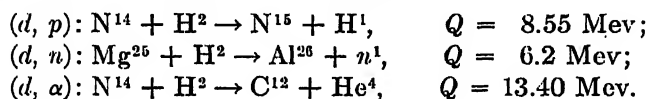
The total number of nuclear reactions known today is well toward 1,000. The list just mentioned contains 685 reactions resulting in induced radioactivity; and many others are known in which the residual nucleus is one of the stable ones that also occur naturally. Out of this wealth of material we can select for detailed discussion only a very few reactions. These we shall group under general statements of the features which they have been chosen to illustrate.² We begin with a few general features that characterize certain classes of reactions.

¹ LIVINGOOD and SEABORG, *Rev. Modern Physics*, vol. 12, p. 30 (1940).

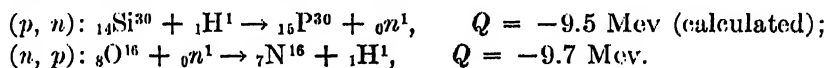
² References, if not given, may be found in LIVINGSTON and BETHE, *Rev. Modern Physics*, vol. 9, p. 245 (1937), or in LIVINGOOD and SEABORG, *loc. cit.*

(a) *Nuclear reactions may be exoergic or endoergic, i.e., Q may be positive or negative.* Q is the net amount of rest energy of particles or nuclei that is converted into other forms; it may also be found by adding to the kinetic energy of the residual nucleus that of any particles that are emitted in the reaction, or the energy $h\nu$ of the γ -ray in a $(-, \gamma)$ reaction, and subtracting the kinetic energy of the bombarding particle and the initial nucleus. To evoke an endoergic reaction, the bombarding particle must possess kinetic energy at least equal to Q ; the actual "threshold" or minimum kinetic energy required may, for other reasons, be higher. The (p, α) reactions discussed in Sec. 219 were all exoergic. Some reactions of this type are endoergic.

Reactions equivalent to the simple addition of a proton or neutron to a nucleus, such as (d, n) , (d, p) , (p, γ) , (n, γ) are nearly always *exoergic*, the nuclear energy decreasing because of the attraction between protons and neutrons that is postulated to explain nuclear binding (Sec. 216). Most (d, α) reactions, also, are exoergic, the net result here being the emission of two proton-neutron pairs in the form of a closely bound α particle as against the capture of a pair in the form of a loosely bound deuteron. For example:



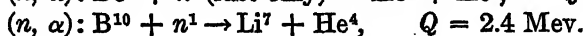
On the other hand, (p, n) or (n, p) reactions are always (?) *endoergic*. A (p, n) reaction merely substitutes a proton for a neutron in the initial nucleus, and an (n, p) reaction substitutes a neutron for a proton, which changes Z by unity without any change in the isotopic number A . In a stable nucleus, however, the division of the A constituent particles into protons and neutrons has been adjusted so as to make the energy as low as possible, or very nearly so [Sec. 216(a)]; hence any change in Z without a change in A is almost sure to increase the energy of the nucleus, at the expense of the kinetic energy of the incident proton or neutron. Examples:



For the reason just stated, the residual nuclei produced in a (p, n) or an (n, p) reaction should be radioactive, tending to revert to the initial value of Z by the emission of a β -ray; and such appears usually to be the case.

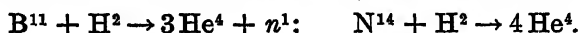
(b) *Neutrons used for bombardment may differ greatly in effectiveness according as they are slow or fast.* Obviously only fast neutrons can

produce an *endoergic* reaction; an *exoergic* one, on the other hand, can be produced by neutrons of any speed, the cross section for such a reaction tending to become very large at low neutron speeds [Sec. 223(f)]. Examples:



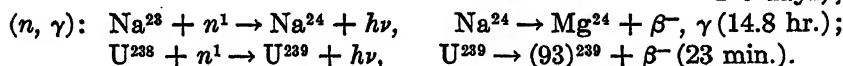
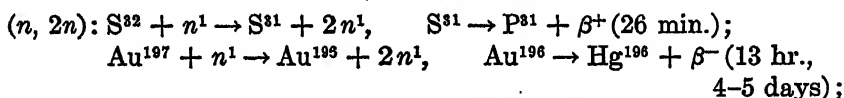
The second of these reactions occurs especially freely with slow neutrons and is much used for their detection [Sec. 223(b)].

(c) *Sometimes three or more particles result from a bombardment, e.g.:*



In these examples, the peculiar stability of the α particle is again in evidence.

(d) *(n, 2n) and (n, γ) reactions produce a residual nucleus isotopic with the initial one, usually detected by means of its radioactivity (Cd being an important exception). Many examples are known, e.g.:*



Here β^+ indicates a positron β -ray and β^- a negative one. The times given are the periods (or times in which half of the radioactive atoms disintegrate). The symbol $h\nu$ is written for a γ -ray that is emitted by the residual nucleus (Na^{24} , U^{239}); this nucleus is left in an excited state and emits the γ -ray in settling into its "normal" state, in which it then remains for a time before undergoing radioactive transformation. The symbol γ is reserved for a γ -ray emitted later by the final nucleus that is formed as the result of radioactive change (e.g., Mg^{24}), this nucleus being left in an excited state by the departure of the β -ray. Thus the third reaction mentioned results in a double γ -ray emission.

The two periods given for Au^{196} represent an example of what is called nuclear *isomerism* (atoms with the same Z and A but different physical properties). Apparently the bombardment produces two different forms of atoms of Au^{196} , which decay radioactively at different rates; perhaps in one form the atoms are produced in a metastable, excited state out of which radiative (or γ -ray) transitions into the normal state happen not to occur.

The properties of the nucleus arising from the decay of U^{239} are still in doubt. Perhaps the element $Z = 93$ may yet be found in nature. (See also Sec. 226 below.)

The reactions just written illustrate two common ways in which the emission of γ -rays may result, directly or indirectly, from the bombardment of nuclei. Several other ways may be mentioned next.

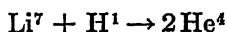
(e) *Reactions of type (α, p) or (α, n) often result directly in the emission of γ -rays; so do some deuteron reactions.* The γ -rays are believed to be emitted by the residual nucleus, which is left in an excited state by the departure of the proton or neutron and emits the γ -ray in descending into its normal state, or into some other state of lower energy. The residual nucleus may or may not undergo subsequent radioactive transformation. An example of such γ -rays, due to a deuteron reaction, is discussed in case (l) below.

On the other hand, (p, α) and (p, n) reactions do not seem to give rise to such γ -rays, the residual nucleus being always left by the reaction in its normal state.

(f) *A nucleus may be merely excited by an elastic collision with a bombarding particle, thereafter emitting a γ -ray as it returns from a higher quantum state to its normal state.* This process, analogous to the excitation of atoms by electron impact in a vacuum tube, might be expected to occur often, but actually it is not common.

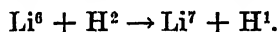
(g) *The intermediate nucleus formed by the capture of the bombarding particle may emit a γ -ray before breaking up.* Such occurrences, however, are rare.

An interesting example of both of these features is presented by one of the first proton reactions to be discovered:



(Sec. 219). This reaction is accompanied by the emission of γ -rays. There appears to be a definite γ -ray line or band at 0.46 Mev, emitted at least whenever the proton energy exceeds 0.85 Mev.¹ These γ -rays have been ascribed to a noncapture excitation of the Li^7 nucleus by collision with a proton, the Li^7 nucleus emitting a γ -ray in returning to its normal state. That is, whereas some Li^7 nuclei capture the H^1 nucleus and then break up into two α particles, according to the reaction just written, others are merely excited as the result of an inelastic collision.

The assumption that the Li^7 nucleus actually possesses an excitation level 0.46 Mev above its normal state is supported by a peculiar feature of another reaction, in which Li^7 nuclei appear as products, viz.,

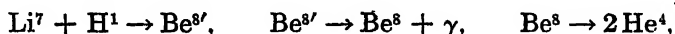


Two groups of protons are produced by this reaction (coming from

¹ HUDSON, HERB, and PLAIN, *Phys. Rev.*, vol. 57, p. 587 (1940).

different individual nuclei, of course); their energy difference, as measured, is 0.45 Mev.¹ It is assumed that the faster protons arise when the Li^7 nucleus is left in its normal state, the slower when it is left in an excited state lying 0.45 Mev higher.

Among the γ -rays from a mass of lithium bombarded by deuterons, there appears also to be a much harder component with a photon energy of about 17.5 Mev.² It is believed that these rays are emitted by an *intermediate* Be^8 nucleus, a complete statement of the reaction first cited being as follows:



the symbol $\text{Be}^{8'}$ standing for an excited beryllium nucleus. The amount of rest mass that disappears in this reaction is:

$$\text{Li}^7 + \text{H}^1 - 2\text{He}^4 = 7.01804 + 1.00812 - 2 \times 4.00388 = \\ 0.01840 \text{ a.m.u.} = 17.15 \text{ Mev.}$$

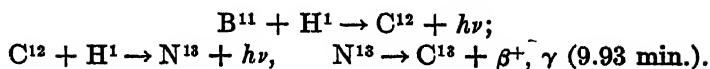
This energy, plus the kinetic energy of the H^1 , must supply both the energy of the γ -ray and the kinetic energy of the two α particles. The kinetic energy of the α particles, by a principle of mechanics, cannot be less than that of the Be^8 nucleus out of which the α particles are formed; and the kinetic energy of Be^8 will be indistinguishable from that of $\text{Be}^{8'}$. The kinetic energy of $\text{Be}^{8'}$, in turn, is $\frac{1}{8}$ of that of the incident H^1 , the $\text{Be}^{8'}$ nucleus being eight times as heavy as H^1 and hence retaining $\frac{1}{8}$ of the velocity of the proton and the same fraction of its kinetic energy. Now resonance, *i.e.*, maximum excitation of the γ -rays in question, occurs when the protons have an energy of 0.44 Mev. Adding $\frac{7}{8}$ of this energy to the energy available from the rest masses, we have a maximum of 17.54 Mev available for the γ -ray. This probably exceeds the energy of the γ -rays, thus leaving something over for the kinetic energy of the α particles, which will be much slower than those due to an immediate disintegration of $\text{Be}^{8'}$ (Sec. 219).

In most instances, on the other hand, the emission of material particles tends to occur so much more quickly than the emission of radiation that γ -rays can be emitted only by nuclei which are either stable or at most α - or β -active.

(*h*) A number of (p, γ) reactions are known, a proton being captured to form a product nucleus in an excited state and the excess energy being then emitted as a γ -ray; the product nucleus may or may not subsequently undergo radioactive change. For example:

¹ RUMBAUGH, ROBERTS, HAFSTAD, *Phys. Rev.*, vol. 54, p. 657 (1940).

² Cf. HUDSON, HERR, and PLAIN, *Phys. Rev.*, *loc. cit.*



Thus, to sum up, γ -rays may be emitted

(1) by the initial nucleus struck by the bombarding particle, which is not captured [Li^7 in case (f)];

(2) by a nucleus formed directly by capture of a bombarding particle [Na^{24} , U^{239} in case (d), C^{12} or N^{13} in case (h)], rarely by such a nucleus which thereafter breaks up by ejecting one or more heavy particles [Be^8 in case (g)];

(3) by the residual nucleus that remains after capture of a bombarding particle and the subsequent emission of one or more heavy particles [case (e)];

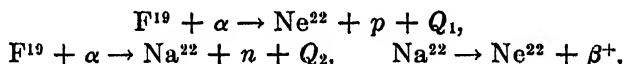
(4) a final nucleus formed by radioactive change [Mg^{24} in case (d)].

(i) *Emitted protons often occur in several groups with different energies.* Besides the example mentioned under (g), another is described below.

(j) *As the energy of the bombarding particle is increased, the frequency of all reactions increases, often more or less exponentially, but with α particles there are often indications of the existence of resonance levels, or energies at which a sudden increase in the yield occurs.*

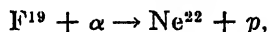
(k) *Sometimes two different reactions result from the same bombardment, some nuclei reacting in one way and others in another.*

These last three features are all well illustrated by the alternative reactions,¹



the letter p standing for a proton, as α does for an α particle. Here the minimum effective α -particle energy, or threshold, is found to be 1 Mev higher for the (α, n) reaction than for the (α, p) , but for α energies well above the minimum the probabilities for the two reactions are of comparable magnitude.

The proton spectrum from the first of these two reactions,



was studied by Chadwick and Constable.² The curve of proton intensity was found to show 6 broad maxima between the extreme proton ranges of 20 cm. and 55 cm. The bombarding α particles came from polonium; their energy could be decreased in any desired ratio by interposing a suitable thickness of absorbing material. When the

¹ SAHA, *Zeits. f. Physik*, vol. 110, p. 473 (1938).

² CHADWICK and CONSTABLE, *Roy. Soc., Proc.*, vol. 135, p. 48 (1932).

energy of the α particles was decreased below that corresponding to an α -particle range of 3.25 cm., the proton spectrum became shortened, one pair of maxima disappearing; at about 2.7 cm., another pair disappeared; and below 2 cm. the whole proton spectrum was absent.

Two distinct causes for the occurrence of such maxima have been recognized and must be carefully distinguished. The maxima may correspond to *different possible states for the residual nucleus*. That is, the emission of the proton may leave the residual nucleus, here Ne^{22} , in any one of several different quantum states. To each of these states will correspond a different value of the reaction energy Q , which represents the (algebraic) excess of the proton and product-nucleus energies over the bombardment energy, or, also, the (algebraic) net loss of rest mass. On the other hand, there is also evidence for the existence of *resonance levels* for the efficiency of the α particles, or energies at which the probability of a transmutation is especially large. Since, for a given final nuclear level, the proton energy must vary equally with the α -particle energy, in order to preserve the energy balance, there will be, for each value of Q , as many maxima in the proton spectrum as there are resonance levels for the α particles, and separated by the same energy differences.

In the observations of Chadwick and Constable, the fact that the proton groups disappeared by pairs as the α -particle energy was lowered indicated that each pair must correspond to a particular resonance energy of the α particles. The respective resonance energies, at which the three pairs appeared with maximum intensity, were estimated at 5.25, 4.0, and 3.4 Mev. The separation of the two groups of protons composing a pair was about the same in all pairs. We can account for the occurrence of the pairs, therefore, if we suppose that the Ne^{22} nucleus may be left in either of two possible quantum states. For reactions leading to these two states, Chadwick and Constable calculated from their data that $Q = 0.99$ or 1.67 Mev, respectively. Thus the two levels of Ne^{22} would lie 0.68 Mev apart (*i.e.*, $1.67 - 0.99$).

That all three pairs were observed simultaneously when high-speed α particles struck the target, in the experiments of Chadwick and Constable, may have been due to the fact that the α particles are slowed down to a stop in the target, so that combined effects due to all speeds below the maximum were observed. It may be that in reality the α resonance levels are fairly sharp, transmutations being almost absent for intermediate energies. If that is so, however, the continuous distribution of proton energies, upon which the low maxima were superposed, must have been due either to unfavorable experi-

mental conditions or to straggling of the protons after they left their point of origin. These questions are not yet cleared up satisfactorily, because of the difficulty of bombarding a target with α particles of sharply defined energy.

The study of the same reaction was extended by May and Vaidyanathan, using the swifter α particles from RaC' (7.0 cm. range).¹ They found three other groups of protons, whose actual ranges as observed were of the same order as those just described, because of the high α energies, but which corresponded to the negative Q values: -3.2 , -2.1 , -0.1 Mev. They observed no duplications resulting from the superposition of various α resonance levels, apparently because they used thin targets and thus kept their α -particle energies within a narrow range. They recalculated Chadwick and Constable's Q values, from better range-energy data for protons, as 0.87 and 1.47. Pose, using polonium particles, had found Q values² which May and Vaidyanathan recalculated as 0.6 and 1.3 Mev. They concluded, therefore, that protons with a Q value of about 1.4 Mev certainly exist, but, finding no trace of an 0.87 or 0.6 group at a point where they should have found it, they suggested that the appearance of such a group in the observations of Chadwick and Constable and of Pose may have been spurious.

If the 1.4 group is assumed to result from the production of Ne^{22} in its normal state, we have then the following energy levels for Ne^{22} , measured from the normal state:

Ne^{22} level.....	0	(0.6)	1.5	3.5	4.6
Q value.....	1.4	(0.8)	-0.1	-2.1	-3.2

Here $1.5 = 1.4 - (-0.1)$, etc. Positive Q means that the protons and the Ne^{22} nucleus together have more kinetic energy than the bombarding α particle; negative, that they have less.

As calculated from the masses, which refer, of course, to the normal nuclear states,

$$Q = \text{F}^{19} + \text{He}^4 - \text{Ne}^{22} - \text{H}^1 = 19.00452 + 4.00388 - 21.99864 - 1.00812 = 0.00164 \text{ a.m.u.} = 1.53 \text{ Mev.}$$

This agrees very well with the observed value of 1.4 Mev.

In direct observations of the *yield* for both the (α, p) and the (α, n) reactions of fluorine as a function of α -particle energy, Saha found indications of the existence of resonance levels for the α particles³:

¹ MAY and VAIDYANATHAN, *Roy. Soc., Proc.*, vol. 155, p. 519 (1936).

² POSE, *Zeits. f. Physik*, vol. 72, p. 528 (1931).

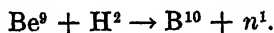
³ SAHA, *Zeits. f. Physik*, vol. 110, p. 473 (1938).

but his values for these levels do not seem to agree at all with those found by Chadwick and Constable!

These fluorine reactions have been discussed at some length, in part, for the purpose of illustrating the general fact that precise work on nuclear phenomena is difficult, and hence conflicting or imperfect data exist in the literature in regard to many nuclear reactions.

(l) *Neutrons, also, tend to be produced in groups, corresponding to different quantum states for the residual nucleus; γ -rays resulting from transitions of the latter nucleus can often be observed.*

Consider the reaction, so much used as a neutron source:



Using 0.9 Mev deuterons (H^2), and determining the neutron energies from recoil proton tracks produced by them in a high-pressure cloud

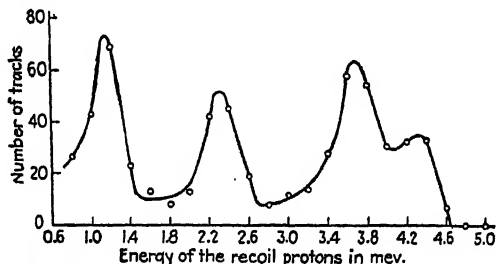


FIG. 212.—Energy distribution of recoil protons due to neutrons from the reaction $\text{Be}^9 + \text{H}^2(0.9 \text{ Mev}) \rightarrow \text{B}^{10} + n$.

chamber containing methane, Bonner and Brubaker obtained the curve shown in Fig. 212 for the recoil protons.¹ Only proton tracks pointing almost straight away from the target were counted, on the assumption that these tracks were made by protons projected in practically the same direction as the incident neutrons.

The four peaks in the curve are interpreted as meaning that the B^{10} nucleus may be left in any one of four possible quantum states. The protons that are projected straight ahead, in collisions with neutrons emitted by deuterons of given energy in a direction at right angles to the deuteron beam, should, therefore, have energies equal to one of four sharply defined values.

Actually, however, it is impossible to screen off high-speed neutrons effectively, so that the walls of the cloud chamber are struck by neutrons over a wide area; many of them, after being deflected by a nucleus in the wall, may, by an oblique collision, produce a proton moving in the direction selected for observation but with reduced

¹ BONNER and BRUBAKER, *Phys. Rev.*, vol. 50, p. 308 (1936).

energy. Furthermore, the deuterons are slowed down somewhat in the target, which causes a corresponding variation in the neutron energy. Thus the proton spectrum may consist, as observed, of four broad peaks instead of four lines, and only the high-speed edge of these peaks represents neutrons projected in the ideal manner, *i.e.*, at 90° by deuterons of maximum energy. Additional small amounts of spreading result, also, from lack of homogeneity in the incident deuteron beam and from straggling of the proton tracks themselves.

From the curve, the authors conclude that neutrons projected at 90° by 0.9 Mev deuterons have energies of 4.52, 4.0, 2.6, or 1.4 Mev. For the neutrons of highest energy we find, from Eq. (312) in Sec. 219, in which $M_1 = 2$, $M_2 = 1$, $M_3 = 10$, nearly enough,

$$Q = ({}^1\frac{1}{10})4.52 - ({}^3\frac{1}{10})0.9 = 4.25 \text{ Mev.}$$

The masses give

$$Q = 9.01497 + 2.01472 - 10.01605 - 1.00893 = 0.00471 \text{ a.m.u.} \\ = 4.39 \text{ Mev.}$$

The last figure of the masses being quite uncertain as given here, the agreement is satisfactory and indicates that in the emission of the 4.25 Mev neutrons the B^{10} nucleus must be left in its lowest state. For the other three groups of neutrons, we find similarly $Q = 3.68$, 2.14, 0.82. Subtracting these numbers from 4.25, we infer energy levels of B^{10} at 0.6, 2.1, and 3.4 Mev above the normal state, about as shown in Fig. 213.

It happens that the γ -rays from this same reaction have been studied by Kruger and Green¹; they confirm the existence of the four energy levels of B^{10} . From the distribution of the tracks of Compton electrons ejected by the γ -rays from a thin piece of mica foil in a cloud chamber, it was inferred that six γ -ray lines occur, with $h\nu$ energies of 0.51, 1.07, 1.44, 1.96, 2.81, 3.21 Mev. These γ -rays could be interpreted as arising from various transitions between four levels of the B^{10} nucleus, as indicated in Fig. 213, where energy levels chosen to fit the γ -ray data are represented by the diagram; the γ -ray ener-

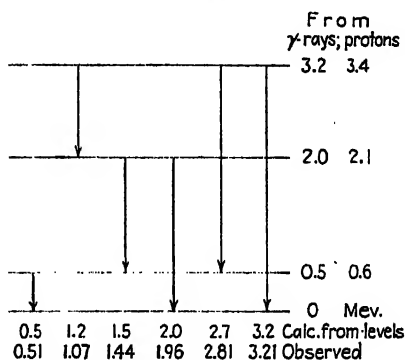


FIG. 213.—Energy levels of B^{10} . All energies, including those of the six γ -rays listed at the bottom of the figure, are in Mev.

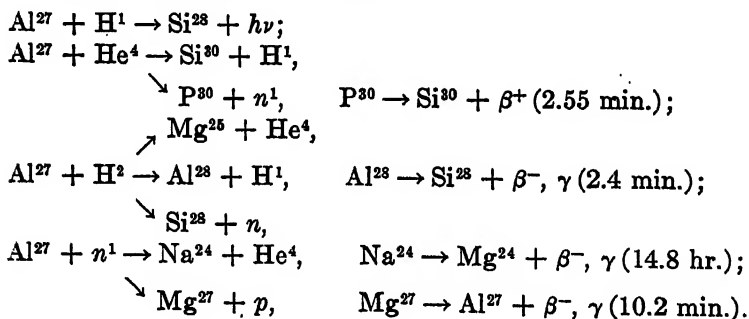
¹ KRUGER and GREEN, *Phys. Rev.*, vol. 52, p. 773 (1937).

gies calculated from these levels are given below the diagram in comparison with the observed energies. The agreement is within the experimental error. The four nuclear levels as determined from the γ -rays agree satisfactorily with those indicated by the neutron groups.

It may be remarked that beryllium bombarded with deuterons also exhibits the following reactions, illustrating again the fact that alternative sets of products often occur:



(m) *The result of a bombardment varies greatly with the nature of the bombarding particle.* This is in harmony with the principle laid down by Bohr, that the bombarding particle is first captured to form an intermediate nucleus, whose composition will depend upon the nature of the captured particle. For example,

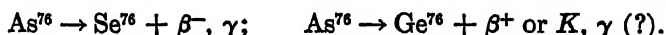


(n) *Artificial radioactive nuclei formed with an increase in $N - P$, or the excess of the number N of neutrons in the nucleus over the number P of protons, tend to exhibit negative β -ray activity; those formed with a decrease in $N - P$, positron activity.* This would be expected as a consequence of the tendency for the ratio N/P to change toward a stable value (Sec. 216). Thus, the principal types of reactions can be classified as follows with respect to the predominant kind of β -ray activity to which the reactions of each type give rise, numerous exceptions occurring except for (p, γ) reactions:

Activity.....	β^-		β^+		
$\Delta(N - P)$	2	1	0	-1	-2
	(n, p)	(α, p) (d, p) (n, α) (n, γ)	(d, α)	(α, n) (d, n) (p, γ) (n, 2n) (γ, n)	(p, n)

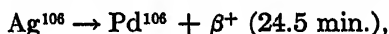
(o) *Radioactive isotopes commonly lie close to stable isotopes of the same element, lighter ones tending to transform radioactively with a decrease of Z , heavier with an increase.* Typical examples are presented by the isotopes of sodium, Na^{22} (radioactive, β^+), Na^{23} (stable), Na^{24} (β^- , γ); and the isotopes of antimony, $\text{Sb}^{115,118,120}$ (all radioactive, β^+), $\text{Sb}^{121,123}$ (both stable), $\text{Sb}^{122,124,127,129,131}$ (all β^-).

(p) *A radioactive isotope lying between 2 stable isobars may be converted sometimes into one isobar, with the emission of a positron, sometimes into the other, with the emission of a negative electron, the period of decay as observed being necessarily the same for both processes.* Example:

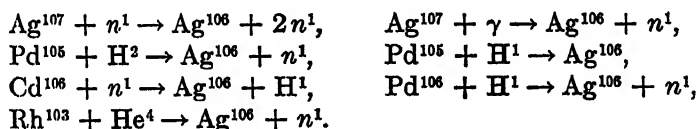


The period is 26.8 hours. The entry, "or K, γ ()," means that there is reason to suspect that capture of a K electron by the nucleus and emission of a γ -ray may also occur as a third alternative, with the same effect on Z as the emission of a positron.

(q) *The same radioactive isotope can usually be produced in more than one way.* As an example, Ag^{106} , with decay reaction,



can be produced in no less than seven ways:

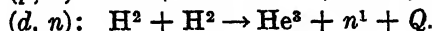
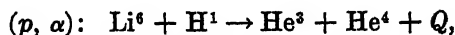


(r) *Radioactive isotopes may be formed of an element otherwise unknown.* At least eight reactions result in the production of radioactive products which should have $Z = 43$ and whose common chemical properties are those to be expected for this missing element. Presumably the absence of this element in nature results from some accidental combination of features that makes every combination of neutrons with just 43 protons unstable.

(s) *The existence in nature of several isotopes of an element complicates the interpretation of observations because of uncertainties as to the particular isotope to which a reaction is to be referred.* In a few cases, however, e.g., lithium and uranium, sufficiently large samples of separated isotopes have been secured so that the assignment can be tested directly.

(t) *Masses of isotopes that do not occur in nature, or at least not in sufficient abundance to be measured otherwise, may often be calculated from a nuclear reaction.* For example, in cyclotron work some indica-

tion has been found for the existence in helium of a light isotope, He^3 , but in extremely minute amounts. Such nuclei should be produced in the two reactions:



The masses of the other nuclei involved here are known. The value of Q for the first reaction was determined by Neuert, and more recently by Perlow, who made very careful measurements of the range of the emitted α particles. The second reaction was studied twice by Bonner; in his second investigation, he measured, by means of recoil protons in methane, the velocities of the neutrons projected straight forward by the deuterons, thus eliminating the effects of angular spread. The date, place, and results of the four investigations are as follows:

	Place	Date	Reaction	Q (Mev)	He^3 mass
Neuert*	Cologne	1935	(p, α)	$3.0172 \pm .0007$
Bonner†	Rice Institute (Texas)	1938	(d, n)	$3.01700 \pm .00010$
Perlow‡	Chicago	1940	(p, α)	3.95	$3.01688 \pm .00011$
Bonner§	Cambridge (England)	1941	(d, n)	3.31	$3.01698 \pm .00006$

* *Phys. Zeits.*, vol. 36, p. 629 (1935).

† *Phys. Rev.*, vol. 53, p. 711 (1938).

‡ *Phys. Rev.*, vol. 58, p. 218 (1940).

§ *Phys. Rev.*, vol. 59, p. 237 (1941).

These results illustrate the increasing accuracy of nuclear work and the close agreement that can be obtained between measurements made in different laboratories.

226. Fission of the Nucleus.—The present chapter dealing with nuclear phenomena will be closed with a short discussion of a very striking phenomenon not discovered until 1939.¹

In 1934, Fermi tried the effect of neutrons upon thorium ($Z = 90$) and uranium ($Z = 92$) and found that β -ray activity was induced in both. The observed decay curve for the activated uranium was such as to point toward the existence of four different radioactive substances, with periods of about 10 sec., 40 sec., 13 min., and 90 min. Fermi found that he could separate the substances exhibiting the 13- and 90-min. activities from the uranium by chemical means, and thus

¹ For a more detailed discussion and further references to the literature see Turner, *Rev. Modern Physics*, vol. 12, p. 1 (1940).

prove that they were not isotopes of that element; and, by further tests, he satisfied himself that these substances were not isotopic with any known element having $Z > 85$. Since it did not seem possible that these substances could have a value of Z extremely different from that of uranium, Fermi concluded that they were new elements with $Z > 92$, which had been produced by one or more processes of β -decay from uranium atoms activated by the capture of a neutron. Thus, the periodic table seemed at last to have been extended upward.

Fermi's results were confirmed and extended, qualitatively, by others. Every experimenter seemed, however, to find a different set of periods. Furthermore, it soon became necessary to assume that the supposed "transuranic" elements ($Z > 92$), in decaying, gave rise to other radioactive elements which would actually be isomeric with known heavy atoms. The number of different substances reported became so numerous and were interrelated in such a way that it was difficult to fit them into any plausible scheme.

Finally, in 1939, Hahn and Strassmann began a thorough investigation in the hope of establishing definitely the chemical nature of the new radioactive substances. In the course of this work, they discovered, in uranium irradiated with neutrons, a new substance with a β -ray period of 250 to 300 hr., which, according to elaborate chemical tests, was either radium ($Z = 88$) or barium (56). To settle the identity of this substance completely, they performed a series of fractional precipitations and crystallizations of the kind used for separating radium from barium. The activity definitely followed the barium, not the radium. After further experiments, Hahn and Strassmann finally felt justified in concluding that the radioactive substance *was* barium. To explain its origin, they proposed the radical hypothesis that *after capturing a neutron, a uranium nucleus may break up into two or more large fragments, each of the size of a moderately heavy atom*. The radioactive barium nuclei they interpreted as fragments produced in this manner by the division of uranium nuclei.

Once the possibility of this new type of disintegration, called by Meitner and Frisch "fission" of the nucleus, was accepted, the facts fell rapidly into order, and fresh corroborations were rapidly obtained. Many of the fission products were recognized as substances already familiar in the study of induced radioactivity. Thorium and protactinium (Pa 91) were also found to undergo fission when bombarded with neutrons. Radioactive isotopes of 18 elements, with Z ranging from 35 to 53, also of Hg (80) and of Bi (83), have been identified among the fission products produced in thorium and uranium,

The energy released in such a fission is enormous. A uranium nucleus (92) might break, for example, into barium (56) and krypton (36). The heaviest known isotopes of these elements have $A = 138$, 86, respectively, or a total of 224, as against 239 for the U^{239} formed by capture of a neutron. Hence, the fragments would have to be new isotopes of great weight, such as Ba^{143} , Kr^{91} , which would then rapidly rise in Z , by successive emission of negative β particles until they attained a stable ratio of neutrons to protons. Now the mass of U^{238} exceeds its own atomic number, $A = 238$, by 0.132 a.m.u., whereas the mass of an ordinary Ba atom is less than A by about 0.084, and that of krypton, by 0.070. If we could assume mass defects of the same magnitude for the heavy Ba and Kr isotopes resulting from fission, we would have available

$$0.132 + 0.084 + 0.070 = 0.286 \text{ a.m.u.} = 267 \text{ Mev of energy.}$$

Some of this energy, at first, would be stored in the fission nuclei as energy of excitation, to be released later in β -ray activity, but much of it should appear at once as kinetic energy of the fission nuclei, which should, therefore, be projected apart at high speed. Huge pulses of ionization ascribed to such fission nuclei have, in fact, been observed in ionization chambers, the pulses indicating total energies of some 150 Mev. These pulses have proved to be a convenient means of detecting the occurrence of the fission process itself.

Variation of the speed of the incident neutrons was found to cause a marked variation in the number of fissions produced in uranium. Fission can be produced either by fast neutrons or by thermal neutrons, but not by neutrons of intermediate speed; the thermal neutrons are effective in the inverse ratio of their speed.

An interesting theoretical problem arose when these facts concerning the production of fission were considered in combination with the following phenomenon. The bombardment of uranium with neutrons has also been found to give rise to a substance, chemically identical with uranium itself, which shows negative β -ray activity, with a period of 23 to 24 minutes, presumably changing into the transuranic element $Z = 93$. The yield curve for this activity shows a resonance maximum at a neutron energy of 25 electron-volts. Now the radioactive uranium nucleus thus produced ought surely to contain more energy than one which has merely captured a thermal neutron. Why, then, does it never undergo fission? Since neutrons of 25 ev energy can excite this β -ray activity, why can they not also cause fission?

As an explanation, Bohr suggested that perhaps the thermal neutrons can cause fission only in the isotope U^{235} , which forms about 0.7 percent of natural uranium, whereas the radioactivity just mentioned results from the reaction, $U^{238} + n^1 \rightarrow U^{239}$. He supposes that fission of a synthetic U^{239} can occur only if the energy brought in by the captured neutron greatly exceeds 25 ev; hence, fission results from the bombardment of U^{238} only if the impinging neutrons are fast ones. Bohr's explanation has recently been shown to be correct by means of experiments upon samples of the separated isotopes. The samples, amounting to 17×10^{-6} gram in the case of U^{238} , were prepared in a mass spectograph.¹ Thermal neutrons cause fission only of U^{235} , with a capture cross section inversely proportional to their speed and of the order of 4×10^{-22} cm.². Fast neutrons are required for fission of U^{238} .

The theoretical explanation of the process of nuclear fission itself presents an interesting problem. Why can it occur only in the three elements of largest atomic number? Perhaps the answer is to be found along the lines of the following picture. We may think of the nucleus as being normally pulled into a spherical shape by the short-range attractive forces between neutrons and protons [Sec. 216(a)], just as a suspended drop of water is made spherical by surface tension. If the nucleus became flattened through some cause, the potential energy due to these forces would be increased. The electrostatic potential energy due to repulsion of the protons would, on the contrary, decrease, but in a lesser degree, so that the equilibrium would be stable. If the deformation became large, however, the decrease in electrostatic energy would probably become predominant, and the deformation would then tend to increase further. If the nucleus actually became pinched apart into two fragments, both types of energy would decrease further as the fragments separated and rounded themselves off into spheres, and the final total energy might thus be less than the energy of the original nucleus. Now we have seen [Sec. 216(a)] that the ratio of electrostatic energy to short-range energy increases with increasing Z . Presumably for $Z = 90$ to 92 this ratio is so large that the nucleus has only a narrow margin of stability against change of shape, and for $Z > 92$ (or 93 ?) equilibrium is entirely impossible. The nucleus U^{235} is perhaps just on the verge of instability, so that it becomes unstable and breaks in two upon the mere addition of one more neutron; whereas to make an unstable U^{239} out of U^{238} the added neutron must also bring in a considerable amount of energy, thereby setting the product nucleus into vibration.

¹ NIER, BOOTH, DUNNING, and GROSSE, *Phys. Rev.*, vol. 58, p. 475 (1940).

A further and very important feature of nuclear fission is the following. Since any heavy atom contains an excess of neutrons as compared with the stable isotopes of its possible fission fragments, we might anticipate that *free neutrons* would also be found among the *products* of fission. In the case of uranium, it is actually found that 2 or 3 neutrons are emitted in each fission process.

Now this phenomenon seems to open up startling possibilities. If each neutron emitted could be made to cause a fresh fission, accompanied by the emission of more neutrons, a "chain reaction" would occur; the number of neutrons and the rate of fission would increase rapidly. Thus a tremendous amount of energy might be released in a fraction of a second. In this way a bit of uranium might perhaps be made to act as a tremendously strong explosive; or, if the process could be controlled, we should have an atomic source of power. The energy released by the fission of all the atoms in a pound of uranium would be 1 million times as great as that obtainable by burning a pound of coal.

Several factors, however, militate against the occurrence of a practically usable chain reaction. The neutrons may be captured in other ways; they will be slowed down by successive elastic impacts with nuclei; and they tend to leave the scene by diffusion. At the moment (1940) it is not known whether or not it is feasible to procure the necessary conditions for a chain reaction.

CHAPTER XII

COSMIC RAYS

The two most active fields of research in physics at the present time are nuclear research and the phenomena attending cosmic rays. Many of the features of cosmic rays lie more properly in the realm of meteorology than of physics, but within the last ten years some of these phenomena have turned out to possess novel physical features of supreme interest. Nature has, in fact, provided in the atmosphere a gigantic laboratory in which atoms are bombarded, not with projectiles of energies limited to a mere 5 or 10 million electron-volts, but with rays of various kinds having energies up to many billions of volts. In this chapter a short account will be given of the more prominent phenomena due to this bombardment.¹

227. Early Work on Cosmic Rays. (a) *Discovery of the Rays.*—It has been known for nearly 100 years that the air possesses a slight electrical conductivity. In 1899–1900 a careful study of the phenomenon was made by Elster and Geitel² and by C. T. R. Wilson.³ The method used was to mount an electroscope on an insulating support such as amber, inside a closed vessel, and to observe the rate at which electrical charge was lost by the electroscope. Wilson took the precaution of connecting the other end of the insulating support to a source of potential equal to the initial potential of the electroscope, so that leakage along the support would tend to maintain the charge and the observed loss of charge could only be due to its neutralization by ions collected out of the air. The conductivity of the air enclosed in the electroscope was found to be permanent, in spite of the continual removal of ions from it by the electrical field. From this fact it may be inferred that the ions are continually being regenerated by some agency.

¹ For further information the student may be referred to F. Rasetti, "Elements of Nuclear Physics," 1936, and to various articles in *Rev. Modern Physics*, vols. 10, 11; also A. H. Compton, *Rev. Sci. Instruments*, vol. 7, p. 71 (1936), and R. A. Millikan, "Cosmic Rays," 1939.

² ELSTER and GEITEL, *Phys. Zeits.*, vol. 1, p. 11 (1899); (GEITEL, *Phys. Zeits.*, vol. 2, p. 116 (1900).

³ C. T. R. WILSON, *Cambridge Phil. Soc., Proc.*, vol. 11, p. 32 (1900),

The cause of this natural ionization of air was generally supposed to lie in traces of radioactive material residing in the apparatus or on its surroundings, or perhaps in the earth or the atmosphere. Other experimenters found that the ionization varied with the material of which the electroscope or ionization chamber was constructed, and that it could be partially screened off by surrounding the apparatus with a heavy layer of water or lead. In the expectation that the ionization should be much less at a considerable height above the earth, Gockel ascended in a balloon and made observations at various heights. He found that the ionization did decrease with increasing height, but only slightly.¹ In 1911–1914 Hess and Kolhörster extended balloon observations² to much higher altitudes and found that above a few hundred meters the conductivity *begins to increase again*. Kolhörster showed that the increase was continuous up to 9,000 meters, at which 93 ions per cm.³ per sec. were produced in the ionization chamber attached to his electroscope, as against 13 at the earth's surface.

To explain the greatly increased ionization at these high altitudes, Hess proposed the novel hypothesis that it was caused by a penetrating radiation falling upon the earth from the outside. He remarked that this radiation could not be coming in large part from the sun, since the ionization was found to be sensibly the same by night as by day. Its penetrating power must be much greater than that of the hardest γ -rays known. For the very hard γ -rays from Th C'', the value of the mass absorption coefficient μ/ρ , or absorption coefficient divided by density (Sec. 195), is 0.036 in Al or 0.041 in Pb, hence perhaps 0.035 in air. Thus a beam of these γ -rays would be reduced to half strength in a thickness x of air at standard density given by $e^{-\mu x} = \frac{1}{2}$, where $\mu/\rho = 0.035$ and $\rho = 0.001293$, or of magnitude $x = 153$ meters. In his balloon observations, Kolhörster found that the ionization decreased to half in a descent of roughly 2,200 meters, at an average height where the density of the atmosphere is $\frac{5}{12}$ of its density at the surface. At normal density this would mean a decrease to half in about 1,000 meters, or $6\frac{1}{2}$ times the distance for the γ -rays.

The observations of Hess and of Kolhörster may be regarded as constituting the discovery of cosmic rays. The latter became known in Germany as "Höhenstrahlung" or "Ultrastrahlung"; the name, cosmic rays, is due to Millikan and Cameron (1925).

¹ GOCKEL, *Phys. Zeits.*, vol. 11, p. 280 (1910); vol. 12, p. 595 (1911).

² HESS, *Phys. Zeits.*, vol. 12, p. 998 (1911); vol. 13, p. 1084 (1912); vol. 14, p. 610 (1913); KOLHÖRSTER, *D. phys. Ges., Verhandlungen*, vol. 15, p. 1111 (1913); vol. 16, p. 719 (1914).

(b) *First Observations of Cosmic Rays under Water.*—An important step forward was taken by Millikan and Cameron¹ when, in 1928, improving upon earlier observations by Kolhörster,² they lowered sealed electroscopes to various depths below the surfaces of snow-fed mountain lakes, which should be especially free from radioactive contamination. The type of instrument used is described in the next section. Even over the water the ionization was observed to be less than over the land, presumably because γ -rays from the earth were screened off by the water. As the electroscope was lowered into the lake, the ionization fell off rapidly in the first meter or so, presumably owing to screening off of γ -rays from the atmosphere. Thereafter it decreased more slowly with increasing depth, and less rapidly as the depth grew greater, in the manner illustrated by Fig. 215 below.

The ionization did not vary as a simple exponential function of the depth, however, as it would if it were caused by homogeneous radiation entering the water vertically from above. In part, the observed variation could be explained by assuming that the incident rays are distributed in direction; for rays entering the water obliquely would pass through more water and so be weakened more in descending to a given depth than would rays entering vertically. Even allowing for this complication, however, the observations indicated that the rays causing the ionization at great depths were much harder (*i.e.*, more penetrating) than those that are responsible for most of the ionization at small depths.

Observations were made in two lakes in California, Arrowhead Lake, at an elevation of 5,100 ft. (latitude 34°), and Gem Lake at 9,080 ft. (latitude 38°). In Gem Lake, the ionization was still decreasing at 50 meters below the surface. The ionization in the lower lake was found to be less than it was at the same depth in the lake at the higher elevation. When, however, the comparison was made between a point at a given depth in the lower lake and a point in the upper one situated farther below the surface by a depth of water equivalent in weight to the column of air between the elevations of the two lakes, close agreement was found. Hence, by plotting cosmic-ray intensities against the total depth below the top of the atmosphere, expressed in equivalent meters of water, the observations in the two lakes could be represented by a single curve. This fact is explained if we assume that (1) equal weights of water and air absorb the rays equally and (2)

¹ MILLIKAN and CAMERON, *Phys. Rev.*, vol. 28, p. 851 (1926); vol. 31, p. 921 (1928).

² KOLHÖRSTER, *Preuss. Akad., Verhandlungen*, 1923, p. 366.

the initial intensity of the rays as they enter the earth's atmosphere is the same in the two localities (which were 250 miles apart).

The part of the combined curve lying just above and just below the equivalent of sea level, treated as approximately exponential, gives a value of 0.25 per meter for the *apparent* absorption coefficient of the cosmic rays in water, or an apparent mass-absorption coefficient $\mu/\rho = 2.5 \times 10^{-3}$ cm.² gram⁻¹. In air ($\rho = 0.001293$) the apparent absorption coefficient would then be 0.32 per kilometer. Such numbers do not represent true absorption coefficients, however, for two reasons: (1) the rays are undoubtedly not homogeneous, and (2) they vary widely in their directions of motion.

These results reinforced so strongly the evidence from the balloon observations that belief in the existence of a penetrating radiation of extraterrestrial origin became general. From this time on the number of workers in the field rapidly increased. Until about 1932, the main object of investigation was the magnitude of the ionization due to cosmic rays as a function of elevation, position on the earth, time of day, and so on. During this period the radiation itself was generally believed to consist of photons resembling γ -rays but of much shorter wave length; and speculation was common as to the possible place of origin of such photons. Apparently they could not be coming from the sun or even from the stars; for the cosmic-ray ionization has always been found to be almost, or quite, uninfluenced by changes in the position of the sun or of the star masses composing the galaxy. Yet it has been estimated that the total amount of energy brought to the earth by cosmic rays is comparable with that brought in by starlight.¹

228. The Measurement of Cosmic-ray Ionization.—Before proceeding with the discussion of cosmic-ray phenomena, it may be useful to discuss briefly the experimental methods employed in measuring the ionization. For this purpose, a string electrometer of very small capacity has usually been used. Millikan and Cameron, in the work discussed in the last section, used an instrument of the Wulf type, the essentials of which are shown diagrammatically in Fig. 214.² The two fibers *FF*, of platinized quartz, hung from an insulating quartz rod, are charged by means of the charging wire *C* and tend to spread apart because of their mutual repulsion. The spreading is resisted by the downward pull of the quartz bow *B*. The positions of the wires are read by means of a microscope (not shown); the wires are kept in a

¹ REGENER, *Zeits. f. Physik*, vol. 80, p. 666 (1933).

² MILLIKAN and OTIS, *Phys. Rev.*, vol. 27, p. 645 (1926).

plane perpendicular to the axis of the microscope by the attraction of induced charges on the metal wires *AA*.

Sometimes only one fiber is employed, opposite an earthed plate. In any case, the position of the fiber is either observed visually from time to time, or it may be projected upon a slowly moving photographic film driven by clockwork, so that the instrument is self-recording. In the latter case, the fiber may either be illuminated intermittently by an attachment connected to the clockwork, or a small spot on it may be made to produce a continuous trace on the film. The clock can also be made to charge the electroscope from time to time. Carefully made self-recording apparatus of this kind has been mounted on ships by Millikan and sent on long voyages over the ocean in order to compare the ionization at different geographical locations. The apparatus required no other attention than the periodic winding of the clock.

In any instrument for measuring ionization there is a zero effect due to residual radioactive contamination of the instrument itself. This zero effect is commonly determined by either sinking the instrument deep under water or taking it down a coal mine, the residual ionization there observed being assumed to be due to radiation from the walls of the instrument. In later instruments, the residual effect was rendered unimportant by using large ionization chambers attached to the electroscope, so as to increase the volume of the enclosed gas in comparison with the area of the walls, and also by raising the pressure of the gas (*e.g.*, to 30 atmospheres). Often other gases than air are used, in which the ionization is greater. Reduction to air at standard density is then made by means of comparison tests using γ -rays from radium. In making measurements near the surface of the earth, the apparatus is commonly shielded with 5 to 10 cm. of iron or lead, in order to cut off radioactive rays from the surroundings.

In comparing ionization at different places, the same instrument is used at all locations. Thus the comparative results do not depend upon the assumption that the ionization per cubic centimeter is the same in instruments of different design.

The *absolute magnitude* of the ionization, or the number of ions formed per cm^3 per sec., is determined by an elementary calculation from the known capacity of the electroscope system and the observed

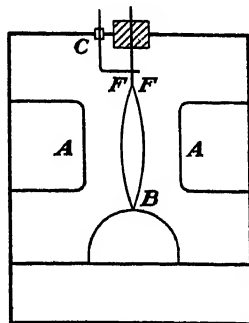


FIG. 214.—The electroscope of Millikan and Otis for measuring cosmic-ray ionization.

rate of decrease in its potential. It has been hard to obtain consistent results on this point. Thus, Millikan¹ obtains a sea-level value of 2.5 ions per cm.² per sec., in standard air, whereas Clay and Jongen² find 1.3, and Compton³ finds 1.2. The cause of these discrepancies is not known. On the other hand, data on the ratios of the ionization at different places, obtained by different observers, agree very well.

229. The Altitude-depth Curve.—Until about 1930, the primary subject of investigation in cosmic-ray work was the variation of the

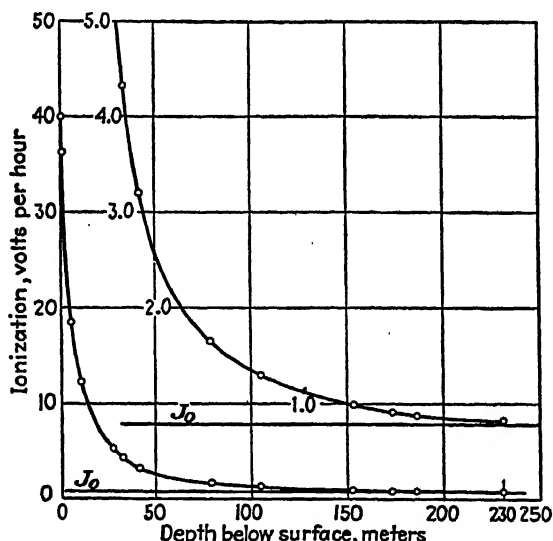


FIG. 215.—Variation of cosmic-ray ionization below the surface of Lake Constance (Regener). The ordinate gives the ionization in terms of volts lost per hour by the electrometer, the scale for the upper curve being 10 times as great as for the lower.

ionization as a function of altitude in the atmosphere or of depth under absorbing material. A few typical results may be cited.

(a) *Observations under Water.*—The most precise observations of cosmic-ray intensity under water are perhaps those of Regener.⁴ In Fig. 215 is shown his curve for the ionization at various depths below the surface of Lake Constance (altitude, 395 meters above sea level, latitude, 47.5° N.). The upper curve repeats part of the lower one on a tenfold greater scale. We note that at 200 meters below the surface

¹ MILLIKAN, *Phys. Rev.*, vol. 39, p. 397 (1931).

² CLAY and JONGEN, *Physica*, vol. 4, p. 245 (1937).

³ COMPTON and TURNER, *Phys. Rev.*, vol. 52, p. 799 (1937); cf. also COMPTON, WOLLAN, and BENNETT, *Rev. Sci. Instruments*, vol. 5, p. 415 (1934).

⁴ REGENER, *Zeits. f. Physik*, vol. 74, p. 433 (1932); cf. also *Phys. Zeits.*, vol. 34, p. 306 (1933).

the ionization, reduced to 2 percent of its surface value, is still decreasing with increasing depth. The value labeled J_0 in the figure was presumed to represent the residual effect in the instrument. If J_0 is subtracted from the ordinates of the curve for the greater depths, the result is a fairly good exponential curve corresponding to a value of the mass-absorption coefficient (in water) of about $\mu/\rho = 0.19 \times 10^{-3}$. This is almost 200 times smaller than the value $\mu/\rho = 0.035$ cited above, in air, for the hard γ -rays from Th C'. At 8 meters below the surface, the apparent value of μ/ρ according to Regener's curve is about 0.76×10^{-3} . At a similar depth (10 meters below the surface at sea level) the data of Millikan and Cameron give about 0.9×10^{-3} . Thus the cosmic rays that cause ionization in deep water must be something like 100 times more penetrating than the hardest γ -rays from radioactive substances.

(b) *Cosmic-ray Intensity at Various Altitudes above the Earth.*—As a typical example illustrating the variation of the ionization with altitude may be cited the results obtained by Millikan, Neher, and Haynes¹ near Fort Sam Houston in Texas (latitude 39°). Instruments were carried up into the stratosphere by sounding balloons, each instrument containing a recording electroscope and a recording barometer and thermometer, all operated by clockwork. The electroscope was recharged every 4 minutes from a charged condenser, which was so well constructed that it lost by leakage only $\frac{1}{2}$ percent of its charge per hour. The instrument weighed 2.5 pounds and was carried aloft by five 1-meter balloons. As the top of the flight was approached, the rate of ascent was diminished by the bursting of two or three balloons; and after $3\frac{1}{2}$ hours the instrument, automatically detached from the balloons, floated back to earth supported by a parachute. The parachute was of silk colored red, so as to attract attention, and an envelope offering a reward for the return of the instrument was attached. Four out of five instruments were eventually recovered, two within 12 hours, one from a distance of 80 miles.

The results from two of these flights are shown by curve C in Fig. 220 (page 664). The ordinate represents the number of ions per cm.³ per sec. that would be produced by the cosmic rays in air at standard density, which serves as a measure of their intensity. The actual ionization at high altitudes is much less because of the lowered density of the air. The abscissa in the figure represents the depth below the top of the atmosphere in equivalent meters of water (calculated as pressure in cm. Hg times $10.33/76$, sea level being, therefore, at 10.33 on the axis of abscissas). The altitude itself can also be calculated

¹ MILLIKAN, NEHER and HAYNES, *Phys. Rev.*, vol. 50, p. 992 (1936).

from the recorded pressure, being about 92,000 ft. for the last point shown on the curve ($p = 1.29$ cm. Hg, 98 percent of the way to the top of the atmosphere).

The *pressure* should be a more significant quantity than the *altitude* for cosmic-ray work, however, since it gives directly the mass of air above the point in question. For this reason—to digress for a moment—the cosmic-ray ionization at a given point ought to vary with the barometric reading; and such is indeed found to be the case. From a long series of readings at different barometric pressures, Compton and Turner deduced¹ a decrease at Chicago of 1.6 percent for each increase of 1 cm. Hg in the barometric height. One might expect to obtain a similar value from the altitude curve. Since 1 cm. Hg means an increase in the effective depth of the atmosphere equivalent to 10.33/76 meter of water, the apparent absorption coefficient of the atmosphere at the earth's surface, as inferred above [Sec. 227(b)] from the observations of Millikan and Cameron, would lead us to expect a decrease per cm. Hg of $0.25 \times 10.33/76 = 3.1$ percent. This agrees only as to order of magnitude with the direct value of 1.6 percent.

Curve *C* in Fig. 220 shows a maximum ionization of 240 ions per cm.² per sec. (in standard air), as against about 2 at the earth's surface. The most interesting feature, however, is the *decrease* within the uppermost 5 percent of the atmosphere. This represents a real decrease in the ionizing power of the cosmic rays. Did the instruments, then, actually ascend above the *source* of the rays, which is, after all, in the atmosphere itself? A more probable explanation is believed to be the following. As the cosmic rays, whatever they are, enter the atmosphere, they produce secondary rays of some sort; the ionizing power of the secondary rays is greater than that of the primary, and hence, as the whole set of rays, primary and secondary, proceeds downward toward the earth, their combined ionizing power increases for a time. Eventually, however, the number of secondaries will reach an equilibrium value, and thereafter the ionization will decrease again, as all rays lose energy in collisions with air molecules or are stopped. What, then, are these mysterious rays that come out of the depths of space? And what is the nature of the secondaries that they produce? Are they photons? High-speed electrons? Protons? Or some new kind of particle, hitherto unknown to physics?

230. Discovery of the Latitude Effect.—In 1927, the Dutch physicist Clay made observations of cosmic-ray intensity during a voyage between the Netherlands and Java (7° S.) and observed that the

¹ COMPTON and TURNER, *Phys. Rev.*, vol. 52, p. 799 (1937).

intensity was distinctly less in equatorial regions than in Europe.¹ This variation, confirmed by him during later voyages and by other observers, is hard to explain on the assumption that the cosmic rays are photons; for there is no obvious reason why photons coming out of the depths of space should tend to avoid the equatorial regions of the earth. Just such a variation is to be expected, however, if the cosmic rays are *high-energy charged particles, perhaps electrons*. For a charged particle would be deflected by the magnetic field of the earth; if it were approaching the polar regions it would be moving more or less parallel to the magnetic lines and would undergo a minimum of deflection, but if its line of approach lay near the equatorial plane it would be deflected to a maximum extent and might be unable to reach the earth at all.

The general theory of the effect of the earth's magnetic field upon the motion of charged particles through it is very complicated. For particles moving in the equatorial plane, however, the theory becomes very simple. Since the principal features of the motion are well illustrated in this special case, the appropriate theory may be developed here as an illustration. For this purpose, as is done also in treating the general case, we shall idealize the earth's field somewhat by supposing it to be exactly the field due to a magnetic dipole of a certain strength located at the center of the earth.

231. Motion of a Charged Particle in the Equatorial Plane of a Magnetic Dipole.—In the equatorial plane of a dipole, which is a plane drawn through the dipole and everywhere perpendicular to its magnetic axis, the magnetic lines are everywhere perpendicular to the plane, and the field strength is

$$H = \frac{M}{r^3} \text{ oersteds,}$$

where M is the magnetic moment of the dipole and r is the distance from its center. In this field a moving particle carrying a charge of e electromagnetic units, positive or negative, experiences a force

$$F = evH = \frac{evM}{r^3}$$

which deflects it sideways; to an observer looking in the direction of the magnetic lines (or *northward* at a point on the earth's equatorial plane) the deflection is *counterclockwise* for a *positive* particle, *clockwise* for a *negative* one. (Cf. Fig. 216, where the plane of the paper represents the equatorial plane, the dipole being indicated by M . The curves

¹ CLAY, *Akad. Amsterdam, Proc.*, vol. 30, p. 1115 (1927).

AA , BB illustrate possible paths which might be followed in the direction of the arrows by positive particles if the magnetic field is directed away from the reader, or by negative ones moving in the opposite direction, or moving with the arrows provided the field is directed toward the reader.) Being perpendicular to the velocity of the particle, the force does no work, and hence *the kinetic energy of the particle and its speed v remain constant.*

Let ψ denote the angle between the radius drawn from the instantaneous position of the particle to the dipole and the tangent to its

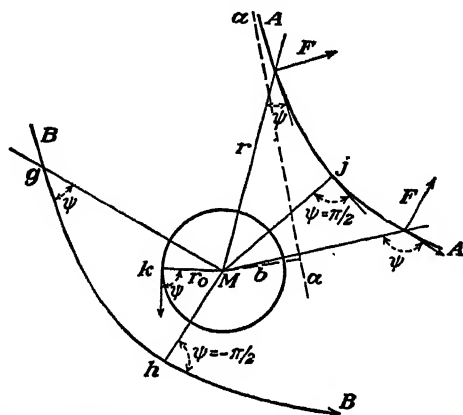


Fig. 216.—Diagram to illustrate the motion of a charged particle in the equatorial plane of a magnetic dipole M .

path, drawn in the direction of motion (Fig. 216); let this angle be measured positively in the usual counterclockwise direction from the former line to the latter (hence, negative in Fig. 216 at g and h). The component of the momentum p of the particle perpendicular to the radius is $p \sin \psi$; and the moment of momentum of the particle about the center M , measured positively clockwise (contrary to the usual convention), is $rp \sin \psi$. (Here at slow speeds we can write $p = mv$, but for energies of cosmic-ray magnitude it is necessary to use the relativistic formula.) The component of the force F perpendicular to the radius, on the other hand, is

$$F \cos \psi = evH \cos \psi = \frac{evM}{r^3} \cos \psi.$$

Hence, by the usual principle connecting moment of momentum and moment of force,

$$\frac{d}{dt}(rp \sin \psi) = rF \cos \psi = \frac{evM}{r^2} \cos \psi.$$

But

$$\frac{dr}{dt} = -v \cos \psi. \quad (316)$$

Hence, by substituting in the preceding equation,

$$\frac{d}{dt}(rp \sin \psi) = -\frac{eM}{r^2} \frac{dr}{dt};$$

and, upon integrating, we have for the moment of momentum,

$$rp \sin \psi = \frac{eM}{r} + pb, \quad (317)$$

where we have written pb arbitrarily for the constant of integration. Thus we can write

$$\sin \psi = \frac{r_0^2}{r^2} + \frac{b}{r}, \quad r_0 = \sqrt{\frac{eM}{p}}. \quad (318a, b)$$

The reason for the choice of notation will appear presently.

Equation (318a) fixes the slope of the trajectory at every point in terms of r . Hence, in general, if one position of the particle is known, the trajectory can be traced. If we introduce a new variable, $\rho = r/r_0$, and a new constant, $\beta = b/r_0$, Eq. (318a) becomes $\sin \psi = \rho^{-2} + \beta \rho^{-1}$. It is thus made evident that a single set of curves for various values of β will serve to represent, in terms of r_0 as the unit of distance, the trajectories for particles of all masses and energies.

Ambiguity arises, however, in tracing a trajectory past a point where $\sin \psi = \pm 1$, because there $\sin \psi$ ceases to vary with ψ , and hence ψ is not given accurately enough by Eq. (318a). Or, in geometrical terms, it is evident that from any point where $\sin \psi = \pm 1$ the path could be continued along a circle about M , with fixed r and $\sin \psi = \pm 1$, thus satisfying Eq. (318a); whereas this may not be the true continuation of the path at all. To see how to get past such points, we may employ the time rate of change of ψ as obtained by differentiating Eq. (318a):

$$\cos \psi \frac{d\psi}{dt} = -\left(\frac{2r_0^2}{r^3} + \frac{b}{r^2}\right) \frac{dr}{dt},$$

whence, by (316),

$$\frac{d\psi}{dt} = v \left(\frac{2r_0^2}{r^3} + \frac{b}{r^2} \right). \quad (319)$$

The algebra does not actually give us this equation at an instant when $\cos \psi = 0$, but the equation holds both before and after such an instant, and all quantities involved in it are continuous functions of the time,

hence it holds also when $\cos \psi = 0$. Thus the equation will tell us how ψ varies along the path near a point where $\sin \psi = \pm 1$.

It is worth noting that, if b is given the proper value, these equations possess a "singular" solution representing a circular orbit about M as center, with radius equal to r_0 . On a circular orbit $\psi = -\pi/2$, $\sin \psi = -1$, as at k in Fig. 216, and $d\psi/dt = 0$; by (318a) and (319), these conditions are satisfied if

$$-1 = \frac{r_0^2}{r^2} + \frac{b}{r}, \quad \frac{2r_0^2}{r^2} + \frac{b}{r^2} = 0$$

or if

$$r = r_0, \quad b = -2r_0. \quad (320a,b)$$

The circular orbit could also be deduced by more elementary methods, of course. We may call it the *critical circle*. Its *radius*, which is the constant r_0 , given by (318b), above, *diminishes with increasing momentum or energy of the particle*.

For the application to cosmic-ray particles, we are interested principally in orbits that extend to infinity. A particle at infinity will be approaching along a certain straight line (such as aa in Fig. 216); and the distance of this line from M , multiplied by p , will represent the initial moment of momentum of the particle about M . If we make $r = \infty$ in Eq. (317), we find for this initial moment of momentum the value pb . Hence, the constant b introduced above represents the distance of the line of approach at infinity from the location of the dipole at M .

The possible orbits now fall into two distinct classes according as $b \geq -2r_0$.

1. $b > -2r_0$.—The properties of these orbits differ according as $b > 0$ or $b < 0$. At $r = \infty$, $\psi = 0$. If $b > 0$, Eq. (318a) shows that $\sin \psi$ increases as r decreases; hence ψ continually increases up to the value, $\psi = \pi/2$, at which point the path is perpendicular to the radius (*e.g.*, at j in Fig. 216). The corresponding value of r , symbol r_1 , represents the distance of closest approach to the dipole, along this particular path, or the distance to the "perigee" of the orbit; its value may be found by putting $r = r_1$ and $\sin \psi = 1$ in Eq. (318a):

$$1 = \frac{r_0^2}{r_1^2} + \frac{b}{r_1}; \quad \therefore r_1 = \frac{1}{2}(b + \sqrt{b^2 + 4r_0^2}). \quad (321a,b)$$

(We reject the other root because it is negative.) At this point $d\psi/dt > 0$, as is evident from (319) and (321a); hence the orbit swings outward again and recedes back to infinity; in fact, it is symmetrical about the perigee. (*Cf.* orbit F in Fig. 217, where, as also in Fig. 216,

the orbits are only very roughly correct.) Since, from (321b), for $b > 0$, $r_1 > r_0$, none of these orbits for $b > 0$ penetrates the critical circle. The orbits for $b = 0$ just touch it, with $r_1 = r_0$ (e.g., *E* in Fig. 217).

If $-2r_0 < b < 0$, on the other hand, the term b/r in Eq. (318a) predominates at first over the term r_0^2/r^2 , and $\sin \psi < 0$. Hence ψ decreases at first (cf. orbits *C*, *D* in Fig. 217), reaching a minimum value when $d\psi/dt = 0$ or, by (319),

$$0 = v \left(\frac{2r_0^2}{r^3} + \frac{b}{r^2} \right), \quad r = -\frac{2r_0^2}{b},$$

and, from (318a),

$$\sin \psi = -\frac{b^2}{4r_0^2}.$$

Since we are at present assuming that $b > -2r_0$, so that $|b| < 2r_0$, this last equation for $\sin \psi$ defines a real negative angle. Thereafter ψ increases again, and the orbit, after crossing the critical circle, passes perigee at $r = r_1$ as given by Eq. (321b) and then returns to infinity. If b is close to $-2r_0$, however, the minimum of ψ is close to $-\pi/2$, and at this point r is close to r_0 , and dr/dt , which equals $-v \cos \psi$, is much smaller than v ; hence the particle may revolve several or many times about the critical circle before plunging inward and reversing its direction of revolution.

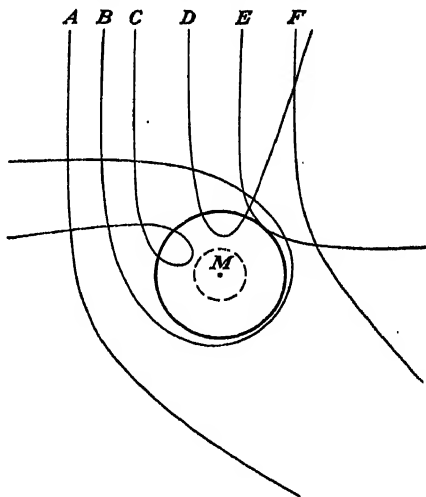


FIG. 217.—A few orbits (only roughly correct) for a charged particle in the equatorial plane of a magnetic dipole *M*.

2. $b < -2r_0$, $|b| > 2r_0$.—In this case r and ψ both decrease until ψ becomes $-\pi/2$, at which point, by (318a), $r = r_2$, where

$$-1 = \frac{r_0^2}{r_2^2} + \frac{b}{r_2}, \quad r_2 = \frac{1}{2}(-b + \sqrt{b^2 - 4r_0^2}).$$

Since $-b > 2r_0$, $r_2 > r_0$. Hence in (319), at $r = r_2$,

$$\frac{2r_0^2}{r_2} < 2r_0, \quad b < -2r_0, \quad \therefore \frac{d\psi}{dt} = \frac{v}{r_2^2} \left(\frac{2r_0^2}{r_2} + b \right) < 0.$$

Thus ψ continues to decrease. The orbit is, in fact, symmetrical about $r = r_2$ as perigee. Since $r_2 > r_0$, all orbits coming from infinity for $b < -2r_0$, like those for $b > 0$, lie entirely outside the critical circle (e.g., orbits A, B in Fig. 217). For these same values of b it can be shown that there exists another set of orbits which lie entirely inside the circle and are periodic; but with these we are not concerned.

For the applications to cosmic-ray particles, the feature of greatest interest about the orbits just described is the *least distance from the center*, r_{\min} , to which the particle can approach for any value whatever of b . This will be the smallest possible value of r_1 for $b > -2r_0$. From (321b), we find that

$$\frac{dr_1}{db} = \frac{1}{2} \left(1 + \frac{b}{\sqrt{b^2 + 4r_0^2}} \right) > 0 \quad (322)$$

for any value of b , since $b/\sqrt{b^2 + 4r_0^2} > -1$ even if $b < 0$. Hence the smallest value of r_1 occurs for the smallest admissible value of b , and, putting $b = -2r_0$ in (321b), we find

$$r_{\min} = (\sqrt{2} - 1)r_0 = 0.414r_0. \quad (323)$$

The actual orbit for $b = -2r_0$, to be sure, does not come so close as this to the dipole; it is a spiral wrapping endlessly about the critical circle and approaching it asymptotically from the outside. But, for $b > -2r_0$, orbits can be found whose perigee distance lies as close as desired to r_{\min} as defined by Eq. (323).

A second point of physical interest is that, as stated above, *all orbits for $-2r_0 < b < 0$ cross the critical circle*; orbits of this type can be found, in fact, *crossing at all possible angles*. For, by (321b), $r_1 = r_0$ at $b = 0$; and hence, in view of (322), $r_1 < r_0$ for any $b < 0$. The value of ψ as the orbit crosses, symbol ψ_0 , is found by putting $r = r_0$ in (318a):

$$\sin \psi_0 = 1 + \frac{b}{r_0}.$$

As b ranges from 0 to $-2r_0$, ψ_0 obviously ranges from $\pi/2$ to $-\pi/2$. At the same time, r_1 ranges from r_0 down to r_{\min} or $(\sqrt{2} - 1)r_0$.

232. Motion of Charged Particles in the Magnetic Field of the Earth.—With the mathematical results of the last section before us, we are prepared to discuss the conditions under which charged particles approaching the earth in its geomagnetic¹ equatorial plane will reach its surface. The results of the general theory will then be stated

¹ The geomagnetic equatorial plane of the earth is a plane perpendicular to the axis of that dipole which, placed at its center, best reproduces its magnetic field.

for any direction of approach, without proof. In all cases the earth's field will be imagined replaced by the most nearly equivalent dipole field.

(a) *Particles Approaching in the Geomagnetic Equatorial Plane.*—If the momentum p of the particles is so small that r_{\min} , as given by (323) and (318b), is greater than the radius a of the earth, the particles cannot reach the surface at all. We have $r_{\min} = a$ for a momentum $p = p_1$, where

$$a = r_{\min} = (\sqrt{2} - 1)r_0 = (\sqrt{2} - 1)\sqrt{\frac{eM}{p_1}}, \quad (324a)$$

$$p_1 = (\sqrt{2} - 1)^2 \frac{eM}{a^2}. \quad (324b)$$

Here M is the earth's dipole moment and e is the numerical charge on the particle, both in electromagnetic units.

As p increases above p_1 , r_{\min} decreases and particles begin to strike the surface at small glancing angles. They will come from the west if they are positively charged, from the east if negatively charged (positive ones moving counterclockwise in Figs. 216 and 217, if the geographic north pole is below the plane of the paper). When p reaches the value

$$p_2 = \frac{eM}{a^2}, \quad (325)$$

r_0 , as given by Eq. (318b), has shrunk to the value $r_0 = a$. All orbits crossing the critical circle are now intercepted by the earth; since we saw above that orbits cross this circle at all angles, the particles may now strike the earth at all angles above the horizon. The same is true at all higher values of the momentum.

If we wish to speak of the velocity v of the particle in place of its momentum p , we may calculate v from the relativistic formula, Eq. (67) in Sec. 67,

$$p = \frac{mv}{(1 - v^2/c^2)^{1/2}},$$

m being the rest mass. The velocity of cosmic-ray particles is nearly that of light, however, and hence varies little. It is more useful to introduce the kinetic energy E . According to Eq. (281) in Sec. 196

$$E = mc^2 \left[\left(1 + \frac{p^2}{m^2 c^2} \right)^{1/2} - 1 \right]. \quad (326)$$

At small energies, i.e., if $p \ll mc$, this reduces to $E = p^2/2m = mv^2/2$. If $p \gg mc$, or $E \gg mc^2$, the approximation

$$E = pc - mc^2 \quad (326a)$$

is useful.¹ The kinetic energies E_{10} and E_{20} , corresponding to the critical momenta p_1 and p_2 at which the particles just begin to reach the earth or to fall upon it at all angles, respectively, can be found by substituting for p in Eq. (326) the values of p_1 and p_2 as given by Eqs. (324b) and (325), viz.:

$$\text{For } E_{10}, \quad p = (\sqrt{2} - 1)^2 \frac{eM}{a^2}; \quad \text{for } E_{20}, \quad p = \frac{eM}{a^2}.$$

If we insert in these formulas, for the earth,²

$$M = 8.04 \times 10^{25} \text{ e.m.u.}, \quad a = 6,370 \text{ km.},$$

and also $c = 2.998 \times 10^{10}$, $e = 4.803 \times 10^{-10}/2.998 \times 10^{10} \text{ e.m.u.}$, $m = 0.9107 \times 10^{-27} \text{ gram}$ for electrons or $m = 1.0076/(6.023 \times 10^{23}) \text{ grams}$ for protons, we obtain:

$$\text{For electrons, } E_{10} = 10.2 \times 10^9 \text{ ev}; \quad E_{20} = 59.4 \times 10^9 \text{ ev.}$$

$$\text{For protons, } E_{10} = 9.3 \times 10^9 \text{ ev}; \quad E_{20} = 58.5 \times 10^9 \text{ ev.}$$

These are enormous energies; and it is noticeable that at such high energies the difference in mass between proton and electron makes very little difference in the deflection of the particles in a magnetic field.

(b) *Particles Approaching from any Direction.*—For the theory of the general motion of electrons in the earth's field, there are available the studies made long ago by Störmer,³ who ascribed the aurora to excitation of the earth's atmosphere by electrons coming from the sun. The application of the theory to cosmic rays was discussed by Epstein,⁴ who pointed out that the variation of inertial mass with velocity does not cause complications, and this application was developed in detail by Lemaitre and Vallarta⁵ and others.⁶ The results of the rather complicated analysis can easily be summarized.

If, as we shall suppose, the cosmic-ray particles of a given energy approach the earth equally in all directions, then, as Lemaitre and Vallarta point out, they remain equally distributed in direction in

¹ Equation (326a) is obtained by expanding thus:

$$\left(1 + \frac{p^2}{mc^2}\right)^{1/2} = \frac{p}{mc} \left(1 + \frac{mc^2}{p^2}\right)^{1/2} = \frac{p}{mc} + \dots$$

² "Smithsonian Physical Tables," 1934.

³ Cf. STÖRMER, *Terr. Mag.*, vol. 22, pp. 23, 97 (1917); vol. 35, p. 193 (1930); also *Nat. Acad. Sci., Proc.*, vol. 17, p. 62 (1931), where references to the original publications are given.

⁴ EPSTEIN, *Nat. Acad. Sci., Proc.*, vol. 16, p. 658 (1930); vol. 17, p. 160 (1931).

⁵ LEMAITRE and VALLARTA, *Phys. Rev.*, vol. 43, p. 87 (1933); vol. 47, p. 434 (1935); vol. 49, p. 719 (1936); vol. 50, p. 493 (1936).

⁶ For references see Johnson, *Rev. Modern Physics*, vol. 10, p. 193 (1938).

spite of their deflection by the magnetic field. If the earth's field is that of a magnetic dipole, there is a family of periodic orbits in which electrons might be trapped permanently, the critical circle in the equatorial plane as described above being one of them. Most of the orbits, however, extend to infinity.¹ Critical energies E_1 and E_2 occur in all cases, corresponding to those described above, but these energies decrease with increasing geomagnetic latitude² λ on the earth. At the poles, $E_1 = E_2 = 0$. That $E_1 = 0$ is evident without calculation; for particles of any energy can easily reach the poles, if the field is exactly that of a dipole—they have only to follow the axial line of force, thus undergoing no deflection. At points on the equator ($\lambda = 0$), the values of E_1 and E_2 are those given as E_{10} and E_{20} above.

At a given point on the earth's surface, in geomagnetic latitude λ , particles with energy less than the value of E_1 proper to that latitude cannot arrive at all. As the energy is raised above E_1 , particles begin to arrive at the top of the atmosphere within a certain cone of directions near the horizon; this cone opens from the west if the particles carry a positive charge, from the east if they are negative. As the energy rises, the cone enlarges. It consists of a principal part and a number of narrow bands; within it the rays arrive with the full intensity with which they approach from the depths of space. When the energy reaches the second critical value, E_2 , the cone fills the whole sky; particles now arrive with equal intensity from all directions, just as if the magnetic field were absent. No further change in the arriving particles then occurs with further increase in the energy.

Or, viewing the distribution of the particles over the earth, we may summarize as follows. Charged particles with energy less than a lower limit E_{10} cannot reach the earth's surface on the geomagnetic equator, although they may reach it elsewhere; $E_{10} = 10.2 \times 10^9$ ev (electron-volts) for electrons, 9.3×10^9 ev for protons. Particles with energy exceeding a second critical value E_{20} arrive at all parts of the earth's surface from all directions and hence with full inten-

¹ Some interesting figures and photographs of electron paths are given by Brüche, *Zeits. f. Physik*, vol. 31, p. 1011 (1930).

² Geomagnetic latitude is measured from a plane perpendicular to the axis of that magnetic dipole, situated at the center of the earth, whose magnetic field most nearly approximates that of the earth. A better approximation can be obtained by displacing the dipole some 200 miles toward the western equatorial Pacific Ocean. The actual observed magnetic poles, where the earth's field at its surface is vertical, differ considerably in position from the locations of the geomagnetic poles, or those points at which the geomagnetic latitude is 90° ; e.g., the observed north magnetic pole is located at 70° N., 96° W., whereas the geomagnetic north pole is at 78° N., 69° W.

sity, just as if the magnetic field were absent; $E_{20} = 59.4$ ev for electrons, 58.5 ev for protons. As we move north or south from the equator, the minimum energy for arrival, E_1 , and the minimum energy for full intensity, E_2 , both decrease; between latitudes 30 and 50°, E_2 decreases especially rapidly, until at magnetic latitudes above 60° even electrons or protons of 2 billion electron-volts will arrive from all directions without diminution by the magnetic field.

Further details concerning the intensity of the arriving particles are furnished by Fig. 218, copied from one of the papers by Lemaître

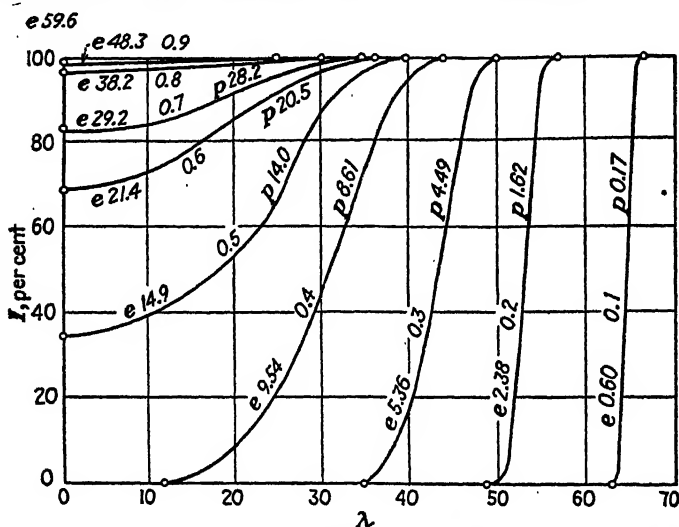


FIG. 218.—Dependence of cosmic-ray intensity on magnetic latitude. (For explanation see text.)

and Vallarta. The abscissa λ represents the geomagnetic latitude, measured from the geomagnetic equator. The ordinate represents the total intensity of the arriving cosmic-ray stream, expressed as percent of the maximum possible intensity; it also represents the fraction of the sky from which the rays come, the number per unit solid angle from this part of the sky being the same as it is outside of the earth's field. The curves drawn in the figure refer to particles having given values of $\rho = a/r_0$, as indicated near the center of each curve [a = radius of the earth, r_0 = radius of the critical circle, defined by Eq. (318b)]; the corresponding energies in terms of billions of electron-volts are given on each curve for two kinds of particles, those for electrons preceded by the letter e and those for protons by p .

In using such curves in the interpretation of the observed ionization due to cosmic rays, however, it must be remembered that rays entering

the earth's atmosphere obliquely have farther to go through it before reaching a given altitude above the earth, and thus will be weakened by absorption much more than are rays that enter vertically. The centers of the curves in Fig. 218, at $I = 50$ percent, should give a fair idea of the latitude at which particles of given energy produce half of their maximum possible effect; but the variation of the effect due to such particles with increase in the latitude will probably occur mostly between the latitudes corresponding to $I = 25$ and $I = 75$ percent.

233. The Variation of Cosmic-ray Intensity with Latitude.¹—In 1930, A. H. Compton, Bennett, and Stearns organized an extensive survey of cosmic-ray intensity at various locations on the earth. Observations were made at 69 stations by a number of cooperating observers, all using similar instruments calibrated on the spot by means of the ionization produced by a standard radium capsule. These observations confirmed the existence of the equatorial drop in intensity, as reported by Clay.

Compton pointed out, however, that his results correlated much better with *geomagnetic* latitude than with *geographic* latitude, and this has been confirmed by later work, especially by the extensive sea-level observations of Millikan and Neher. The latter observers, and Clay independently, discovered in 1934 that even the slight variation in the strength of the earth's magnetic field along the equator is reflected in the cosmic-ray intensity, which is some 4 percent lower in the East Indies and the Indian Ocean, where the magnetic field is strongest, than it is on the opposite side of the earth, in equatorial South America, near which the field is weakest.

All observations indicate, on the other hand, that the cosmic-ray intensity does not vary much from one location to another above geomagnetic latitude 40° , N. or S., which means in the United States above, roughly, 30° in geographic latitude. From a geomagnetic latitude of 40° N. or 40° S. to the equator, the drop amounts to about 10 percent on the average, rising to 12 percent in eastern Asia and sinking to 8 percent in the longitude of equatorial South America. The drop begins rather suddenly as the limiting latitude is passed.

The most accurate study of the equatorial drop is probably that made by Compton and Turner. They mounted a recording electro-scope on a steamship of the Canadian Australasian Steamship Company plying regularly between Vancouver and Auckland (New

¹ Cf. COMPTON, *Phys. Rev.*, vol. 43, p. 387 (1933); MILLIKAN and NEHER, *Phys. Rev.*, vol. 47, p. 205 (1935), vol. 50, p. 15 (1936); COMPTON and TURNER, *Phys. Rev.*, vol. 52, p. 799 (1937). Other references are given by JOHNSON, *Rev. Modern Physics*, vol. 10, p. 193 (1938).

Zealand) and Sydney (Australia), the route crossing the geomagnetic equator almost where it crosses the geographic equator. The observations were extended over 10 months.

A decided *seasonal effect* was noted in these observations, especially at the extreme north and south ends of the range, where the cosmic-ray intensity during the warm months averaged less than during the cold months. In Fig. 219 are shown the curves obtained by Compton and Turner for the four seasons, ordinates representing average cosmic-ray intensity and abscissas the geomagnetic latitude. The seasonal

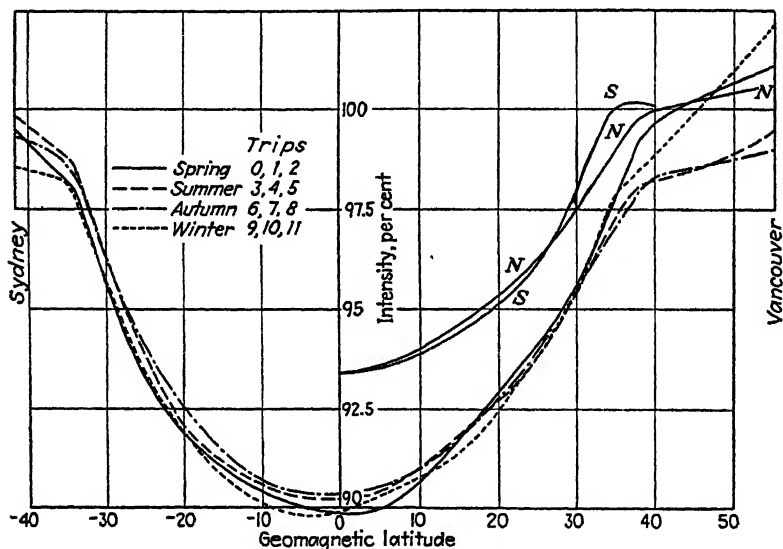


FIG. 219.—Variation of cosmic-ray ionization with latitude on the Pacific Ocean. The value called "100 per cent" is chosen arbitrarily (and differently for the curves *NN*, *SS*). (See text.)

variation, which has also been recorded by others, is believed to be correlated with atmospheric temperature; such an interpretation is supported by observations in Europe showing a variation of the same order of magnitude with temperature at a given station. (Temperature effects on the apparatus itself were shown not to occur.) From their data, Compton and Turner inferred a decrease of about 0.18 percent per degree centigrade of rise of temperature.

The temperature effect was believed by Compton and Turner to account for perhaps 3 percent of the observed decrease of 10 percent from high latitudes to the equator. Correcting for it, they found as the true geomagnetic effect the curve labeled *NS* in Fig. 219, representing ionization corrected to a temperature of 21°C. In drawing

this curve, the southern half *SS* has been folded back over the northern half *NNN* in order to show how nearly they agree.

The close correspondence thus revealed between variations in cosmic-ray intensity and the features of the magnetic field at the surface of the earth constitutes convincing evidence that at least part of the primary cosmic rays must consist, not of photons, but of charged particles of some sort. The effect of the magnetic field upon the particles must be produced before they enter the earth's atmosphere; for the atmosphere is so thin relative to the earth's dimensions that any effect produced within it could have little influence on the large-scale distribution of cosmic-ray intensity over the earth. If all of the primary cosmic rays are charged particles, however, those which cause the ionization near the equator must have energies exceeding 35 billion electron-volts (or $\rho = 0.75$), for otherwise, as is evident from Fig. 218, the latitude effect would be larger than it actually is. An alternative possibility, of course, would be that the equatorial effect is due to an agency (photons?) which, for some reason or other, is little affected by the earth's field, whereas the latitude-sensitive part of the cosmic-ray ionization is due to charged particles of much lower energy.

From the curves of Fig. 218 and the remarks made above concerning the use of these curves, it appears that particles of about $\rho = 0.37$, or electrons of 8 billion or protons of 7 billion electron-volt energy, would show little increase in their effect above latitude 40° . They should also be ineffective below 25° , however, so that the further drop in the ionization from that point to the equator must be due to the presence of particles of still higher energy. Thus we may conclude that the primary cosmic rays, whatever their nature otherwise, must include charged particles of energies ranging upward from 7 or 8 billion electron-volts.

At higher altitudes than sea level, the latitude effect is much greater. As an illustration, altitude curves obtained in different latitudes by Bowen, Millikan, and Neher, with the use of sounding balloons, are shown in Fig. 220. The abscissa represents depth below the atmosphere in meters of water (1 atmosphere = 10.3 meters). We note that the maximum ionization decreases by 73 percent in going from high latitudes to the equator (as against a decrease of 8 to 12 percent at sea level). The decrease is about one half if the comparison is made at altitudes a third of the way down from the top of the atmosphere. The critical geomagnetic latitude, above which variation with latitude ceases, seems to be 50 to 55° in the upper half of the atmosphere, as against 40° at sea level.

The higher value of the critical latitude at higher altitudes indicates the presence of slower particles than those that are responsible for the sea-level effect. Not much energy can be coming in carried by particles of less than 2 billion ev ($\rho = 0.2 \pm$ in Fig. 218), however, otherwise the critical latitude would be higher still. The energies must, therefore, range widely upward from 2×10^9 electron-volts.

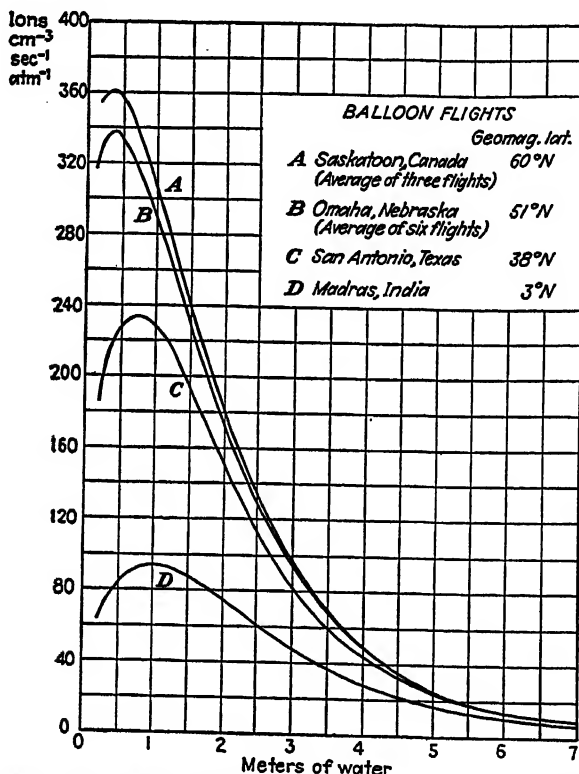


FIG. 220.—Cosmic-ray ionization as a function of altitude in four different latitudes (Bowen, Millikan, and Neher.)

A quantitative analysis made by Johnson¹ points toward a broad energy distribution with a maximum around 10 billion volts and reaching well up toward 50 billion. The number of particles, however, measured per unit of the energy axis, increases continuously with decreasing energy. The curve of total incoming energy falls off below 10 billion volts because of the smaller energy carried per particle. Such calculations, however, must be regarded as essentially tentative.

¹ JOHNSON, *Rev. Modern Physics*, vol. 10, p. 193 (1938), especially p. 236.

234. Observations on Single Cosmic-ray Particles.—The investigations described in preceding sections had to do with the average ionization produced by cosmic rays. Methods have also been invented for the observation of single particles, and such observations might be expected to furnish more direct evidence concerning the nature of the rays. Even if some of the primary rays are high-energy photons, the ionization resulting from them will be nearly all produced by Compton electrons ejected by the photons from the walls of the apparatus or from molecules of the air, and it should be possible to observe these secondary electrons.

Cloud-chamber observations of cosmic-ray particles were first made, accidentally, by Skobelzyn in 1927.¹ (For a description of cloud chambers see Sec. 213.) While studying the secondary β -rays produced by γ -rays from radioactive material, he noticed occasional tracks which were hardly curved at all in his magnetic field of 1,500 gauss. The ionization along these tracks was about as dense as along the tracks of the β -rays; therefore, he concluded that they also were made by electrons. From the smallness of the curvature of the tracks he calculated that the energy of the particles must be above 15×10^6 ev, which exceeds any energies encountered in radioactivity. In direction the tracks were strongly concentrated toward the vertical. Skobelzyn concluded, therefore, that these tracks were those of the particles that are responsible for the ionization ascribed to cosmic rays.

From the frequency of the occurrence of such tracks in his apparatus, Skobelzyn estimated that about 1.2 per minute crossed each square centimeter of a horizontal plane. Assuming a specific ionization of 40 ion pairs per cm. of track, he concluded that the particles would cause a total ionization of $1.2 \times \frac{40}{60} = 0.8$ pairs of ions per second in each cubic centimeter of air. This is of the same order of magnitude as the observed cosmic-ray ionization, 1.5 to 2 ion pairs per cm.³ per sec. A better modern estimate would be an average of perhaps 70 ion pairs per cm. of track [Sec. 235(b)] and 1.7 particles per cm.² per minute; these data give about 2 ion pairs per cm.³ per sec.

Counter observations constitute a second method for the detection of individual charged particles. Whenever a counter is used for any purpose, a slow background rate of counting is always evident. Part of this is doubtless due to radioactive contamination in the walls of the counter, for the background count varies with the material of the wall; but part of it is undoubtedly caused by cosmic-ray particles.

In order to eliminate effects due to rays from the walls of the counters, which are not very penetrating, Bothe and Kolhörster

¹ SKOBELZYN. *Zeits. f. Physik*, vol. 43, p. 354 (1927); vol. 54, p. 686 (1929).

introduced the device of using *two counters* and recording only instances in which both counters discharged at the same moment, which they called *coincidences*.¹ In their work the counter discharges were recorded on a moving photographic film running 1 cm. per sec., and a coincidence was considered to have occurred when two counts were recorded within 0.01 sec. of each other. It will occasionally happen, to be sure, that a coincidence occurs owing to the independent passage of two particles, one through each counter; but the number of such occurrences to be expected in a given time can be calculated from the total counting rates of the individual counters. The number of chance coincidences so determined was subtracted from the observed number as a correction.

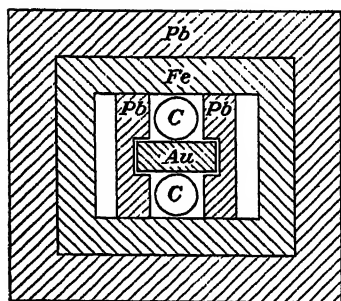


FIG. 221.—Apparatus of Bothe and Kolhörster for measuring the absorption of cosmic-ray particles in gold.

radiations and thereby permit only the highly penetrating cosmic-ray particles to activate the counters. Two blocks *Pb*, *Pb* of lead were placed close alongside the counters in an effort to stop particles that might pass through the upper one obliquely and then be scattered into the lower.

When observations were made with this apparatus near the roof of the building, about 3 coincidences per minute were observed, and this number decreased by 25 percent when the gold absorber was put in place. From this result, Bothe and Kolhörster deduced the value, $\mu/\rho = 3.5 \times 10^{-3} \text{ cm.}^2 \text{ gram}^{-1}$, for the mass absorption of the cosmic-ray particles. A similar experiment performed later by Rossi,² in which the cosmic rays were filtered through a layer of lead 5 cm. thick placed above the counters, gave $\mu/\rho = 1.6 \times 10^{-3}$ in a lead absorber 9.7 cm. thick. For comparison, we may cite the value deduced from Millikan and Cameron's curve for the (apparent) mass absorption coefficient of the

¹ BOTHE and KOLHÖRSTER, *Zeits. f. Physik*, vol. 56, p. 751 (1929).

² ROSSI, *Zeits. f. Physik*, vol. 68, p. 64 (1931).

cosmic rays in air at sea level, $\mu/\rho = 2.5 \times 10^{-3}$. This is of the same order of magnitude as that found for the particles in the counter observations. The value deduced from the depth curve should be higher, as it actually is, partly because few of the rays were filtered out in making the measurements of the ionization, and partly because many of them would be traveling obliquely and so should decrease relatively more rapidly as the thickness of the absorbing layer of atmosphere is increased.

The counter method was improved later by Rossi in two respects. First, he replaced the photographic film by an electromechanical device for recording the discharges of the counters.¹ This was accomplished by connecting each counter to the grid of a three-electrode tube in such a way that the discharge of a counter momentarily lowers the potential of the grid and stops the current flowing through it. The tubes are connected in parallel and function as comparatively low resistances in a circuit that contains also a detector tube. When passage of electricity through the counter tubes is completely stopped by their functioning simultaneously, the grid of the detector tube rises momentarily in potential and so causes a pulse of current in its anode circuit, which causes in turn an audible click in a telephone, or may be made to actuate a mechanical counter of some sort. Such a counter circuit can be used to record coincidences between any number of counters.

The second improvement introduced by Rossi was the use of *three* counters in a row,² an observation being recorded only when all three counters respond simultaneously (or at least within a certain very small fraction of a second of each other). In this way coincidences due to secondary electrons ejected from the surroundings are effectively eliminated. When a primary particle passes through the first counter, only a secondary produced by it in the walls of this counter itself could pass through the other two counters; but the counter walls are very thin, and hence such secondaries should be very few in number. Equally rare should be the occasions when two secondaries are produced in the surroundings by the same primary particle and one of them passes through each of the other two counters. Two secondaries produced by a high-speed photon will also seldom cause a coincidence. A single secondary passing through the entire row of counters, however, could not be distinguished from a primary particle. Thus the arrange-

¹ Rossi, *Nature*, vol. 125, p. 636 (1930). A good circuit for use with a single counter is described by Neher and Harper, *Phys. Rev.*, vol. 49, p. 940 (1936).

² Rossi, *Zeits. f. Physik*, vol. 82, p. 151 (1933). Cf. also TUBE, *Phys. Rev.*, vol. 35, p. 651 (1930).

ment selects charged particles moving within a narrow solid angle of directions. Such an arrangement of counters is sometimes called a *counter telescope*.

Rossi also avoided screening the apparatus closely with lead all round, because, as his investigations showed, secondary particles are produced in such screens in large numbers and may seriously falsify the results.

Using three counters (Fig. 222) in a vertical line, Rossi¹ studied the absorption in lead of cosmic-ray particles which had already been filtered through 7 cm. of lead (placed above the counters). Moving one counter out of line was found to reduce the frequency of coincidences to that which was to be expected as a result of chance (1 in 26 hours as against 1 to 2 per hour with the counters in line). He found that 46 ± 5 percent of the coincidences remained when 101 cm. of lead were placed between the counters, concluding that nearly half of the particles can penetrate such a thickness of lead. If we assume for purposes of calculation that the absorption is exponential through each successive layer of lead, we find for the mean mass absorption coefficient: in the first 10 cm. of lead, $\mu/\rho = 1.8 \times 10^{-3}$ (total absorption, 19 percent); in the next 15 cm., $\mu/\rho = 0.5 \times 10^{-3}$; and in the last 76 cm.,

$$\mu/\rho = 0.55 \times 10^{-3}.$$

FIG. 222.—A "counter telescope" used by Rossi in studying the absorbability of cosmic-ray particles.

These figures show unmistakably that the cosmic rays include charged particles of enormous penetrating power.

The two methods of observing cosmic-ray particles that have been described were brought into relation with each other in 1931 by Mott-Smith and Locher, who placed a cloud chamber between two counters and noted that tracks in the cloud chamber and coincident responses of the counters occurred together. This indicated that the two effects must be due to the same type of charged particle. In 1932, Blackett and Occhialini introduced the system, now generally employed, of allowing the coincident discharge of the counters to release the expansion of the cloud chamber. Such an arrangement is sketched in Fig. 223. The counters can be connected so that a particle passing through any chosen combination of them will release the mechanism that withdraws the piston and operates the cameras. In this way, much useless photographing is avoided (otherwise, it requires 10 to

¹ *Ibid.*

20 random expansions to obtain a track), and only tracks having a suitable position and direction are photographed.

From 1930 on, many investigations of cosmic-ray particles were made by various observers. The interpretation of the results of such experiments presents special difficulties because of the enormous penetrating power of the particles and because of their powerful effects upon matter, and it is often hard to harmonize the results of different investigations. We shall discuss only certain points on which general agreement seems to have been reached, referring the student for further information to recent summarizing articles, such as appear from time to time in the *Review of Modern Physics* and elsewhere,¹ or to the literature.

The interpretation of recent work has often depended upon a precise knowledge of the behavior of high-speed particles. It will be convenient to summarize at this point what is known in regard to certain of their properties.

235. Energy, Mass, and Specific Ionization of Charged Particles. (a) *Measurement of Charge, Mass, and Energy.*—The standard method of crossed electric and magnetic fields (Sec. 44) yields, for a charged particle, its velocity and the ratio of its charge to its mass. Additional observations of some sort are needed in order to find e and m separately; from m and the velocity, the energy can then be calculated.

When the velocity of the particle is almost equal to the velocity of light, however, the velocity becomes difficult to measure with sufficient accuracy to make the measurement significant. There is then no point in using an electric field to deflect the particle; the easier measurement with a magnetic field yields all the information that can readily be obtained.

Formulas for the rapid motion of a charged particle in a uniform magnetic field were obtained in Sec. 196. If M is its mass, q its numerical charge in *electrostatic units*, and ρ the radius of the circle which it describes in a magnetic field of H oersteds, then, from Eqs.

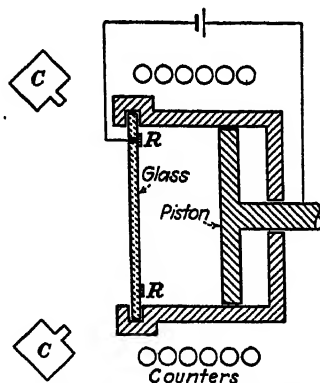


FIG. 223.—A counter-controlled cloud chamber, with stereoscopic cameras CC . An electric field is momentarily applied between the ring R and the piston to sweep the space clear of ions. [After Stearns and Froman, *Am. Phys. Teacher*, vol. 7, p. 79 (1939).]

¹ Cf. *Am. Phys. Teacher*, vol. 7, p. 79.

(279) and (282), the momentum p and kinetic energy E of the particle are given by the equations¹

$$p = \frac{qH\rho}{c}, \quad H\rho = \frac{cp}{q}, \quad (327a,b)$$

$$E = Mc^2 \left[\left(1 + \frac{q^2 H^2 \rho^2}{M^2 c^4} \right)^{1/2} - 1 \right]. \quad (328)$$

If $E \ll Mc^2$, expansion gives, since $(1+x)^{1/2} = 1 + x/2 - \dots$,

$$E \ll Mc^2: \quad E = \frac{q^2}{2Mc^2} H^2 \rho^2 - \dots \quad (328a)$$

If we write E_{ev} for E expressed in electron-volts and m for the electronic mass, we have from the last equation, for a particle of mass M , numerical charge equal to that of the electron, and kinetic energy $E \ll Mc^2$, approximately,

$$E_{ev} = 0.088 \frac{m}{M} (H\rho)^2, \quad (328b)$$

within 1 percent if $E < Mc^2/50$. If $E \gg Mc^2$, we have

$$E \gg Mc^2: \quad E = qH\rho - Mc^2 + \dots, \quad (328c)$$

or, if q equals the electronic charge,

$$E_{ev} = 300H\rho - (Mc^2)_{ev}, \quad (328d)$$

within 1 percent if $E > 10Mc^2$. For the intermediate region, the plot shown in Fig. 224 may be of use in passing from one to another of the three quantities velocity, $H\rho$, and E . Scales, logarithmic in both directions, are shown for both electrons and protons.

An important conclusion from these formulas is that the curvature of the path of a charged particle with energy greatly exceeding its rest energy ($E \gg Mc^2$) is almost independent of its mass ($\rho = E/qH$ nearly). An approximate estimate of the *energy* of such a particle can be made, therefore, from the curvature of its track in a cloud chamber without knowing its mass. Partly for this reason and partly because the particles were long supposed always to be electrons, the "energies" of cosmic-ray particles as determined from cloud-chamber observations are commonly calculated on the assumption of an electronic mass. An energy so calculated might be called an "equivalent electron energy." A more precise method of reporting the data, sometimes employed, is to report the values of $H\rho$.

¹ Calculations from such formulas are most easily made in the following way: $E = Mc^2 [(\cos \tan^{-1} qH\rho/mc^2)^{-1} - 1]$.

The *curvature* of the tracks in a cloud chamber thus gives primarily a measure of the *energy*, or, more precisely, the *momentum* of the particle. The *mass* must then be inferred in some other way. The principal source of evidence as to the mass, and also as to the *magnitude* of the *charge*, has been the amount of ionization caused by the particle in the gas in the cloud chamber. The *sign* of the charge can be inferred from the direction of curvature of the path, provided the direction of motion of the particle is known. In the case of cosmic-ray particles,

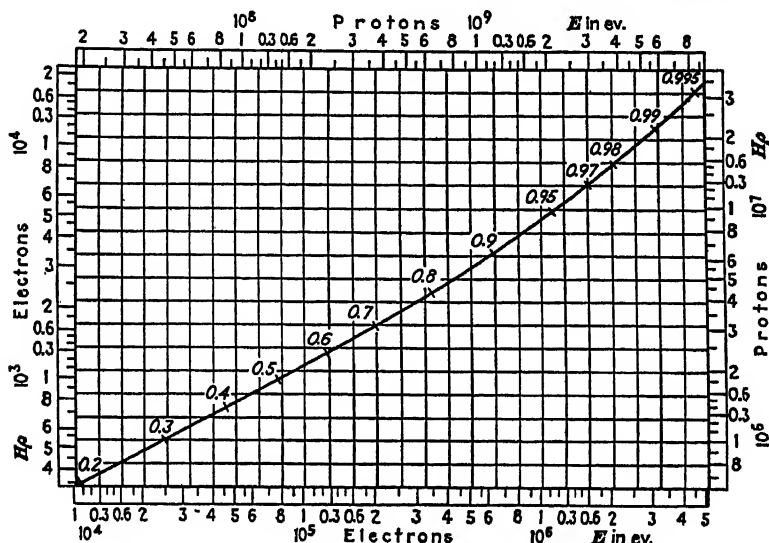


FIG. 224.—Relation between kinetic energy E in electron-volts and $H\rho$ (ρ = radius of curvature in centimeters in a field of H oersteds), for electrons and for protons, on logarithmic scales. Numbers on curve refer to the cross lines and give ratio of velocity of particle to that of light. For a particle k times as heavy as an electron, multiply $H\rho$ and E for an electron by k ; or, multiply by 100 and read on the scales n cross-section spaces to the right of or higher than the point on the curve where for $k = 126, 158, 200, 251, 316$, respectively, $n = 1, 2, 3, 4, 5$.

it may reasonably be assumed that tracks near the vertical were traversed in a downward direction. Or, if a track passes through a slab of absorbing material, it may safely be assumed, as in Anderson's discovery of the positron, that the particle traveled toward the side on which greater curvature of the track gives evidence of lower energy.

(b) *The Specific Ionization in Air.*—The ionization produced by electrons in air has been studied experimentally over a wide range of energies.¹ The ionization produced by a cathode ray has been found to increase with increasing energy of the ray, up to a maximum of several

¹ See E. RUTHERFORD, CHADWICK, and ELLIS, "Radiations from Radioactive Substances," pp. 444ff., 1930.

thousand ion pairs per centimeter of path at an energy of 1,000 volts or so, above which it decreases again. It was found by W. Wilson¹ that for β -rays of energies from 6×10^4 to 1.7×10^6 ev the total ionization produced in standard air was almost inversely proportional to the square of the speed of the β particle, decreasing from about 214 ion pairs per centimeter of path at 6×10^4 ev to about 46 ion pairs at 1.7×10^6 ev.

The ions measured in these observations include those produced by secondary electrons ejected from molecules by the β particles with sufficient energy to produce additional ions. The *primary* ionization produced by the β particle itself can be determined separately by counting the drops of water produced in a cloud chamber (C. T. R. Wilson). The path of most of the secondary electrons is so short that usually all of the ions formed directly or indirectly as the result of a single primary ionization lie at first very close together, and, if the expansion is performed immediately after the passage of the β -ray, a single droplet of water is condensed on the entire cluster. The primary ionization determined in this way is about half as great as the total ionization, at electron energies around 10^6 ev. The *total ionization* itself can also be determined in the cloud chamber by delaying the expansion a fraction of a second (*e.g.*, 0.2 sec.) so as to give time for the ions to spread out by diffusion. If the expansion is great enough, each ion then gives rise to a separate droplet; or, the expansion can be so adjusted that droplets form only on the positive ions.

As far as the appearance of tracks in the cloud chamber is concerned, however, a complication is presented by the fact that occasionally a secondary of exceptionally high energy makes what appears to be a branch track. The appearance of the tracks is determined by the average amount of ionization exclusive of that on such branch tracks; this is called the *probable* ionization.

The term "specific ionization" is applied in a general way to any number representing ions or ion pairs per centimeter.

Approximate theoretical formulas for the ionization produced by charged particles of high energy have been derived from wave mechanics. The way in which the ionization should vary with the energy of the particle is illustrated by certain of the curves in Fig. 233 below [Sec. 237(c)]. Curves *e-ec* and *ep-ep* illustrate the total and the probable ionization, respectively, produced by an electron, whereas the curves labeled *p* and *m160* illustrate the total ionization produced, respectively, by a proton and by a "mesotron" (Sec. 238) 160 times as heavy as an electron, the medium being in all cases air of standard

¹ W. WILSON, *Roy. Soc., Proc.*, vol. 85, p. 240 (1911).

density. Actually the quantity plotted is the rate of loss of energy per cm., but the number of ion pairs produced is nearly proportional to this, being about equal to the energy loss in ev per cm. divided by 32 (for air). References are given under the figure. The theoretical formulas seem to agree with observation, but they have not been checked thoroughly for electrons.

As shown by the curves in Fig. 233, the theoretical formulas predict a decrease in the ionization, due to the decrease in the time during which the particle can act on an atom, as its speed increases, up to approximately the speed of light, when the energy of the particle becomes of the order of its rest energy. Above this point the ionization should increase again as the energy of the particle increases, but very slowly, about as the logarithm of the energy; from 10^8 to 10^{10} electron-volts, the probable ionization should about double. The increase results from the increasing contraction of the electric field of the particle toward a plane perpendicular to its motion (Sec. 36), which increases the sharpness of the impulse given to atomic electrons by particles passing at some distance from the atomic center. A further factor, causing an increase in that part of the total ionization which is produced by secondaries of high energy, is the increase in the maximum energy which can be transferred to an atomic electron in a close collision.

Observations by Corson and Brode¹ on cosmic-ray particles, presumably electrons, with energies up to 30 Mev, gave evidence of the predicted increase. They found a (probable) specific ionization of about 50 ion pairs per cm. of path in standard air at the minimum (just above 10^6 ev). The total ionization will then be somewhat greater. The theoretical formula indicates a minimum of about 50 ion pairs per cm. for the probable and 65 for the total ionization.

The most important feature of the theoretical formulas in their application to cosmic-ray observations is the way in which the ionization should vary with the kind of particle. The ionization should be roughly *proportional to the square of the charge* and, hence, should be about the same for positive and negative particles. Otherwise it should depend, to a close approximation, *only on the velocity*. Since, at given velocity, kinetic energy and mass are proportional to each other, even in relativistic mechanics, it follows that *particles of equal charge but different masses should produce about the same ionization at kinetic energies proportional to their masses*.

¹ CORSON and BRODE, *Phys. Rev.* 53, p. 773 (1936). For a correction see Brode, *Rev. Modern Physics*, vol. 11, p. 222 (1939), especially p. 228.

Because of this feature, corresponding curves for particles of different mass in Fig. 233 are almost exactly alike in shape but displaced relatively to each other along the axis representing the logarithm of the energy. Thus a particle k times as heavy as an electron should produce about the same ionization as an electron with an energy $1/k$ times as great. Herein lies the reason for the relatively heavy ionization ordinarily produced by a proton. According to the theory, however, a proton track, though clearly distinguishable from an electron track up to a proton energy of $10^5 \times 1,837$ or about 2×10^8 ev, should actually be less dense at proton energies exceeding 2×10^9 electron-volts.

Since we have seen that the curvature of the tracks in a cloud chamber gives at once an approximate value of the energy of the particle, it is clear that, over a considerable range of energies, the density of the ionization along the track furnishes information in regard to the mass, at least provided the magnitude of the charge can be inferred in some way with sufficient certainty.

In *different materials*, theory indicates, with some support by experiment, that the total ionization should be roughly proportional to the total number of atomic electrons per cubic centimeter.

236. Showers and Bursts. (a) *Showers*.—When cosmic-ray tracks are observed in a cloud chamber, it frequently happens that two or

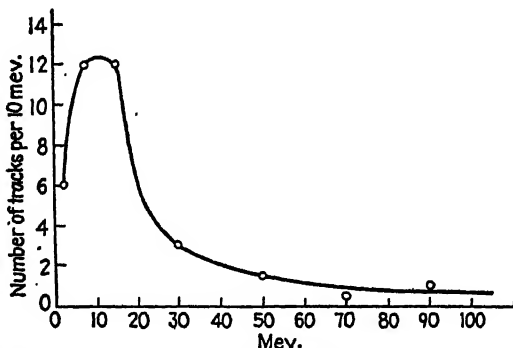


FIG. 225.—Energy distribution of positive and negative electrons occurring as showers in a cloud chamber. (Anderson and Neddermeyer.)

more tracks are seen which appear to have been produced simultaneously; often they diverge from a common center lying in the walls or outside of the chamber. The first observation of this sort was made by Skobelzyn, in the course of the work described above (1927; Sec. 234). If there is a solid obstacle in the chamber itself, such as a sheet of lead, groups of tracks may often be seen diverging



FIG. 226.—Reproduction of the first cloud-chamber photograph showing a track recognized as that of a positron (Sec. 221). [*Photograph kindly lent by Professor C. D. Anderson. See Phys. Rev., vol. 43, p. 491 (1933).*]

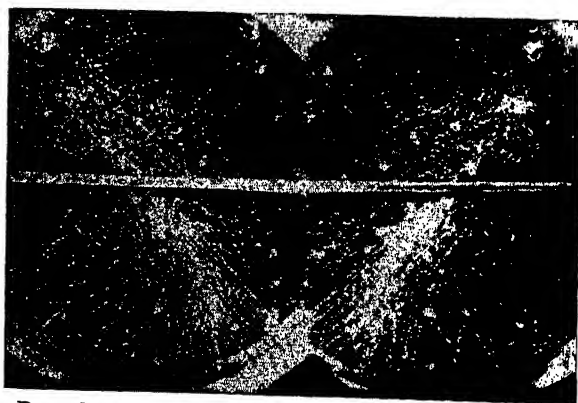
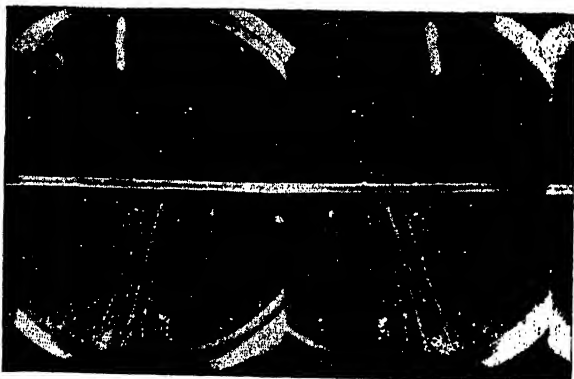


FIG. 227.—Reproduction of three stereoscopic cloud-chamber photographs showing showers, two of them exhibiting multiplication in a lead plate. [Photographs kindly lent by Professor C. D. Anderson. See Anderson and Neddermeyer, *Phys. Rev.*, vol. 50, p. 263 (1936).]

from a point in this obstacle; frequently a single cosmic-ray track coming from above seems to end at the initial point of a group of tracks. (Cf. Fig. 227.)

Such groups of tracks are believed to be made by showers of (negative) electrons and positrons created as the result of the impact of a high-speed particle or photon upon matter. From a statistical study of the curvature of the tracks it is inferred that positive and negative particles occur about equally often, and with energies ranging from 1 Mev up to 500 Mev, 5 to 20 Mev being the most common energy. The distribution curve obtained for them by Anderson and Neddermeyer is shown in Fig. 225.¹

Interesting statistical data on the occurrence of *showers of various sizes* were also obtained by Anderson and Neddermeyer.² They used a counter-controlled cloud chamber divided midway by a horizontal lead plate 3.5 mm. thick, placed in a horizontal magnetic field of 7,900 oersteds and viewed horizontally. At Pasadena, near sea level, out of 2,684 photographs of cosmic-ray tracks, 383 or 14 percent showed showers of two or more particles. On Pikes Peak (elevation, 4,300 meters above sea level), out of 1,775 photographs, 752 or 42 percent showed showers. Two-particle showers were the commonest; but in both places the next commonest size was a shower of 6 to 10 particles. One photograph obtained on Pikes Peak showed more than 300 tracks of electrons and positrons, their total energy being estimated to exceed 15,000 Mev; four others showed showers of at least 100 particles. Often it was evident that the number of particles in a shower increased as the shower passed downward through the lead sheet. Three of their photographs are reproduced in Fig. 227.

The *increased frequency of showers at higher elevations* is an important feature of these observations. An increase in the frequency of all tracks would be expected, since the cosmic-ray ionization itself increases with altitude; but the showers increase much faster than the single particles, especially the larger ones. Frequencies of occurrence were calculated by dividing the number of photographs by the time taken to obtain them, exclusive of the 15-second intervals required after each expansion to reset the cloud chamber. The rate was thus found to be 34 photographs showing tracks per hour in Pasadena and 120 per hour on Pikes Peak. The ratio of the number of photographs per hour showing n tracks on Pikes Peak to the corresponding number at Pasadena, for various values of n , was calculated to be as follows:

¹ "International Conference on Physics," p. 174, University Press (John Wilson & Son, Inc.), Cambridge, Mass., 1935.

² ANDERSON and NEDDERMEYER, *Phys. Rev.*, vol. 50, p. 263 (1936).

n	1	2	3	4	5	6-10	11-20	21-100
Ratio.....	2.7	7.5	6.9	17	24	18	33	24

The increase in the frequency of large showers at the higher elevation is very striking.

(b) *Bursts of Ionization*.—In 1927, the German physicist Hoffmann, while studying the ionization caused by cosmic rays in an ionization chamber, noticed that occasionally a large deflection of his electroscope occurred.¹ He convinced himself that these deflections were not caused by some defect in the apparatus and indicated, therefore, the sudden production of a large “burst” of ionization. The phenomenon has been studied extensively, especially since 1932. Bursts are

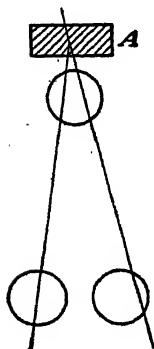


FIG. 228.—Sketch of three noncollinear counters for detection of showers generated in A.

observed to vary greatly in size, their frequency decreasing with increasing size. Apparently no limit can be set as yet to the possible size of a burst. Using an ionization chamber 36 cm. in diameter filled with argon at 50 atmospheres and shielded by 12 cm. of lead, at Huancayo, Peru (altitude, 3,350 meters), Schein and Gill observed a burst of some 10^9 ion pairs, which they ascribed to about 10,000 cosmic-ray particles having a total energy of some 10^{12} ev; and they say that even larger bursts occurred.²

No further details concerning observations on bursts will be given here, however, chiefly because it seems to be well established that they are caused by the same showers of particles that are observed in cloud chambers. Burst observations thus constitute an alternative method of studying showers.³

(c) *Observations on the Production of Showers*.—Still a third method of observing showers of charged particles is by means of three or more counters *not* placed in a row (if placed in line, they count both showers and single particles).⁴ In 1933, Rossi used the arrangement shown in Fig. 228. A coincident discharge of the three counters can only be caused by a shower which contains at least two particles, including the incident one (or, very rarely, by two independent particles passing through the apparatus almost simultaneously).

¹ Cf. “International Conference on Physics,” p. 226.

² SCHEIN and GILL, *Rev. Modern Physics*, vol. 11, p. 267 (1939).

³ For a summary of burst observations see Froman and Stearns, *Rev. Modern Physics*, vol. 10, p. 133 (1938).

⁴ ROSSI, *Zeits. f. Physik*, vol. 82, p. 151 (1933).

With the apparatus unshielded, Rossi observed 6.75 triple coincidences per hour. The number of chance coincidences, calculated from the rates of the counters when counting separately, he estimated at 3 per hour. A lead plate 1 cm. thick, placed above the upper counter at A, increased the coincidences to 14.1 per hour, owing to the production of showers in the lead plate.

The number of showers as a function of the thickness of the shower-producing matter was studied, using a similar arrangement, by varying the thickness of the block A. When the counting rate with the block A removed was subtracted as a correction, the number of the remaining

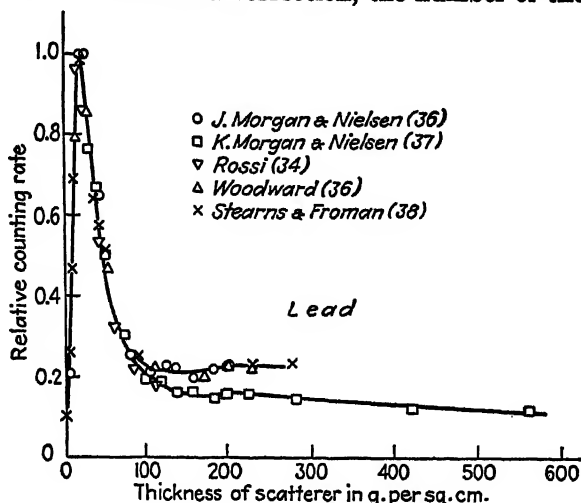


FIG. 229.—Relative numbers of showers observed under various thicknesses of lead.

coincidences, plotted against the thickness of the lead block A as abscissa, gave a curve like the left-hand third of that shown in Fig. 229. Many similar observations have been made subsequently by others, the curve being thereby extended to greater thicknesses. The figure shows the data obtained by several observers. A considerable discrepancy is to be noted for the greater thicknesses.

The maximum number of showers in lead occurs at a thickness of about 1.3 cm. Some observers find a second slight maximum at about 20 cm. Similar results have been obtained using iron or aluminum. Such curves, or similar ones in which the whole number of showers is plotted, are called (Rossi) "transition" curves.

It seems clear that the showers observed in this way are somehow produced in the lead. That their number should *increase at first* with increasing thickness of the lead is reasonable. The *subsequent decrease* in their number, however, as the thickness is increased further,

can only be due to absorption of the primary shower-producing agency. For an increase in the thickness is equivalent to the addition of a fresh layer of lead on top. This added layer will (1) itself give rise to shower particles, some of which may penetrate the lower layers of lead, and (2) absorb some of the shower-producing particles, so that fewer showers will be generated in the underlying lead. If the lead already in place is thick enough so that few of the shower particles produced in the added layer can get through it, the absorption effect will predominate; the number of showers observed below the lead will then be decreased by an increase in its total thickness.

The initial rapid decrease of the showers up to 10 cm. of lead, followed by a very slow continued decrease, must mean, then, that the shower-producing agency, whether it be particles or photons, consists roughly of two parts, a more highly absorbable or "soft" component and a very penetrating or "hard" component. The rate of the slow decrease, corresponding to

$$\mu/\rho = 0.6 \times 10^{-3},^*$$

agrees with the value $\mu/\rho = 0.55 \times 10^{-3}$ found by Rossi for the absorption in lead of cosmic-ray particles filtered through 25 cm. of lead, which would certainly represent the penetrating component. The more absorbable rays which, in the absorption measurements, were removed by 10 to 20 cm. of lead would then represent the soft component.

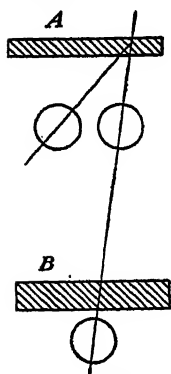


FIG. 230.—Arrangement for measuring the absorptivity of shower particles.

In order to explain the occurrence of a maximum in the transition curve at less than 2 cm. of lead, we must suppose that the shower particles themselves are easily absorbed. This hypothesis can be tested directly by measuring the *penetrating power of the shower particles*. This is easily done by placing an absorbing screen above the lowest of the three counters (e.g., at B in Fig. 230). It is found that the shower particles are mostly absorbed by a few centimeters of lead.

In view of these results, it is a reasonable hypothesis that the soft component of the primary cosmic rays consists of shower particles that have been produced, directly or indirectly, by the penetrating component in the overlying atmosphere or in other material bodies. The mode of initiation and growth of the showers in matter thus becomes a theoretical problem of great importance. We shall devote the next section to a discussion of this problem.

* MORGAN and NIELSEN, *Phys. Rev.*, vol. 51, p. 689 (1937).

237. Theory of the Shower Phenomenon. (a) *The Shower Process.* Until 1934, the origin of showers of cosmic-ray particles was generally held to be a mystery. The materials for a theoretical understanding of them had already been accumulated, however, and during the next three years an apparently satisfactory theory for the common type of shower was developed.

Several physicists, noting that shower tracks usually do not all diverge from a common point, suggested that they may originate in cascade fashion.¹ In 1935, Auger pointed to the production of electron-positron pairs by photons, which was described in Sec. 221, as a possible step in the process. These ideas led to the following picture of the life history of a shower. A high-energy photon becomes converted, in the field of a nucleus, into an electron and a positron [cf. Fig. 231(a)]. These particles constitute a shower of two rays; or, if the photon itself had been produced by a high-speed particle, we have at this stage a shower of three rays. The particles, accelerated in the fields of other nuclei, then radiate part of their energy in the form of fresh photons, this process constituting the extension to high energies of the process by which the continuous X-ray spectrum is produced [Fig. 231(b)]. These photons in turn disappear in giving rise to additional pairs of electrons and positrons; these produce fresh photons; and so on. At such high energies, it can be shown that the particles and photons should usually move almost in the same direction as the ray by which they are produced, so that the whole group of particles and photons will move on in close array.

The number of photons and particles in the shower may continue to increase for some distance. Eventually, however, the energy of a particle becomes so low that it loses more energy in the production of ions than in the production of photons; its remaining energy is then rapidly frittered away and the particle comes to rest. Similarly, the photons come finally to lose energy principally by the ejection of Compton electrons from the atoms, and soon after that they become too weak to produce pairs. Because of these factors, the number of rays (particles and photons) in a shower increases to a maximum and then decreases again. In the end the positrons that have been produced all combine with atomic electrons, and the total number of charged particles in the world is thereby restored to its initial value.

¹ Cf. MONTGOMERY, *Phys. Rev.*, vol. 45, p. 62 (1934).

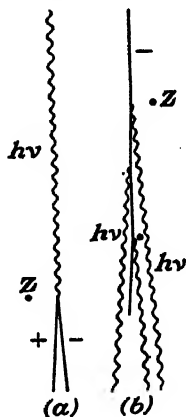


FIG. 231.—Diagrams illustrating the processes by which showers are built up.

The entire history of the shower occupies, of course, only a minute fraction of a second, since the particles are moving almost with the speed of light.

To develop this hypothesis concerning the growth and decay of showers in quantitative form, it is necessary to know the probabilities of the various elementary processes that are involved. We shall consider these in turn.

(b) *The Laws of Pair Production by Photons.*—The theory of the conversion of photons into pairs is given in Heitler's book on radiation.¹ The probability for the occurrence of such a conversion can con-

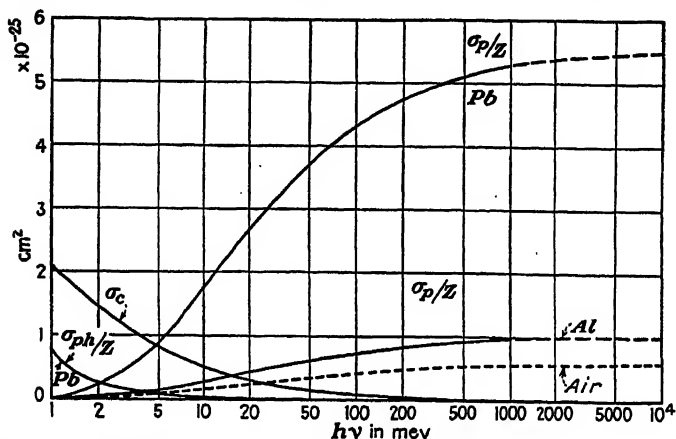


FIG. 232.—Approximate plot of the theoretical cross sections for the action of high-energy photons upon atoms of lead, aluminum, or air: σ_p/Z is the cross section σ_p for absorption of the photon with production of an electron-positron pair divided by the atomic number Z of the atom ($Z = 7.26$ for air); σ_e , the mean cross section for the Compton ejection of each electron from any atom (the cross section for the entire atom being $Z\sigma_e$); σ_{ph}/Z , the cross section for photoelectric absorption of the photon divided by Z (shown for lead only, inappreciable in aluminum or air for $h\nu > 1$ Mev). The photon energy $h\nu$ is plotted logarithmically. (From data in Heitler, "Quantum Theory of Radiation," 1936.)

veniently be stated in terms of a cross section σ_p for pair production. If a uniformly distributed beam of photons, each of energy $h\nu$, passes through matter containing n nuclei per cm^3 , as many photons will be converted into pairs per centimeter of matter traversed as cross an area $n\sigma_p$ perpendicular to the beam. Or, the mean distance L that a photon goes before being converted into a pair is $L = 1/(n\sigma_p)$. Theory indicates that σ_p should vary rapidly with the atomic number Z of the nuclei, in fact almost as Z^2 , especially below $h\nu = 25$ Mev. The theoretical values of σ_p for three materials, divided by Z , are shown in Fig. 232. For comparison, the cross section per electron for Compton

¹ HEITLER, W., "Quantum Theory of Radiation," 1936.

scattering, σ_s , is also shown ($nZ\sigma_s$ being, therefore, the number of Compton electrons produced per photon per centimeter of path), and also the cross section σ_{ph} for photoelectric absorption of a photon in lead, in which photoelectric absorption is still appreciable up to 5 Mev. Some values of L in centimeters are as follows:

Energy $h\nu$, Mev	In standard air	In Al	In Pb
25	9.8×10^4	27.4	1.25
100	5.9×10^4	17.0	.86
1,000	4.5×10^4	13.2	.70

When a photon turns into a pair, part of its energy becomes the proper energy of the electron and the positron that are created; this part amounts to $2mc^2$ or 1.02 Mev. The remainder takes the form of kinetic energy. The latter may be divided in any ratio between the electron and the positron; and the probability for either particle to receive a given fraction of the available kinetic energy is, very roughly, the same for all fractions.

(c) *The Loss of Energy by Charged Particles.*—A charged particle passing through matter may lose energy by any one of the following processes (at least):

- (α) ionization and excitation of atoms;
- (β) emission of radiation due to acceleration in the atomic fields;
- (γ) collisions with nuclei resulting in nuclear transmutations.

In the case of particles of very high energy, (γ) is known to occur but is rare. The principal causes of energy loss are therefore (α) and (β).

The loss due to ionization and excitation can be calculated from the observed ionization, provided we know the mean loss of energy per ion produced. Experiments with β -rays indicate that in air the mean loss is about 32 electron-volts per ion pair, not varying much with the speed of the electron. The observed minimum value of the "probable" ionization, 50 ion pairs per cm., thus implies an energy loss of at least $50 \times 32 = 1,600$ ev per cm. of path in standard air.

Theoretical formulas for the energy loss have been developed and appear to be in agreement with such experimental data as exist. Some curves plotted from the theoretical formulas are shown in Fig. 233, as is explained further below, or under the figure. The theoretical minimum loss due to ionization and excitation by electrons is about 11 Mev per cm. in lead, or 2,100 ev per cm. in standard air, at an electron energy of about 10^6 ev; of the minimum loss in air, about 1,680 ev per cm. corresponds to probable ionization. The curves

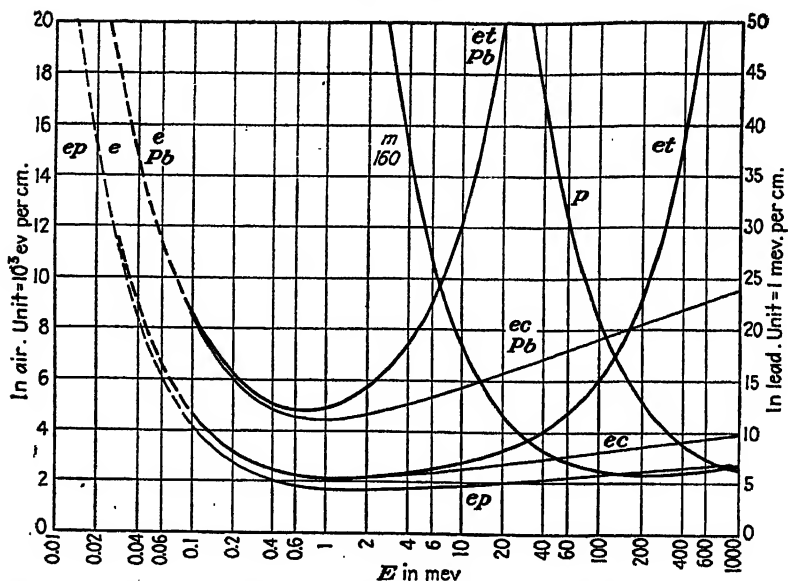


FIG. 233.—Some curves showing the approximate energy loss by charged particles passing through matter, as calculated from wave mechanics. Abscissa: kinetic energy E of the particle on a logarithmic scale. Ordinate: expectation $(-dE/dx)$ of loss of energy by a particle per cm. (N particles all of energy E lose $N(-dE/dx)dx$ units of energy in going a distance dx). Curve $ePb - ecPb$: loss due to ionization and excitation of atoms by an electron in lead. $ePb - etPb$: total loss when the direct emission of radiation (due to acceleration of the incident electron) is included, the difference in ordinates between this curve and the preceding representing the loss due to radiation alone. Curve $ep - ep$: "probable" loss in standard air by an electron due to ionization and excitation, omitting single collisions in which more than 10^4 ev is lost. $e - ec$, the same loss when all collisions are included, the fastest electron that leaves the scene of an ionizing collision being regarded as the incident one. $e - et$, loss by the same electron when the direct emission of radiation is included. Curves $m160$ and p : total loss in standard air by a mesotron 160 times as heavy as an electron, and by a proton, respectively, the radiation loss being negligible. Broken parts of curves are less certain.*

* For mesotrons and protons, and for the "probable" loss by an electron, Bloch's formula [*Zeits. f. Physik*, vol. 81, p. 363 (1933)] was used, as modified by Babba for particles of spin $\frac{1}{2}$ [*Roy. Soc., Proc.*, vol. 164, p. 257 (1938)], viz.: $(-dE/dx) = (2\pi e^4 NZ / \mu \beta^2) \{ \log [2\mu \beta^2 W / (1 - \beta^2) I^2 Z^2] - 2\beta^2 \}$, where e = electronic charge, N = number of nuclei per unit volume and Z is their atomic number ($Z = 7.26$ for air), $\mu = mc^2$ = rest energy of an electron, β = ratio of velocity of the incident particle to that of light, $I = 13$ ev, W = maximum allowed loss in a collision [$W = 10^4$ ev for the "probable" loss, otherwise $W = 2\mu(2M + E)E / [(M + \mu)^2 + 2\mu E]$ where M is the rest energy and E the kinetic energy of the incident particle]. For electron collisions with all losses included, Møller's expression [*Ann. d. Physik*, vol. 14, p. 531 (1932)] was integrated up to $W = E/2$, giving

$$(-dE/dx) = (2\pi e^4 NZ / \mu \beta^2) \{ \log [\mu \beta^2 E / (1 - \beta^2) I^2 Z^2] + 1 - \beta^2 - [2 - (1 - 1/\gamma)^2] \log 2 + (1 - 1/\gamma)^2 / 8 \},$$

$\gamma = (1 - \beta^2)^{-1/2}$; below $E = 0.1$ Mev, however, the curve in Fig. 233 was merely continued so as to join the probable-loss curve at 10^4 ev.

for heavier singly charged particles, such as protons, are nearly the same but with all energies increased in the ratio of the mass. Thus the minimum loss for a proton should occur at an energy of about 1,800 Mev.

The energy lost in radiation, owing to acceleration of the charged particle in the fields of the nuclei (partly screened by the atomic electrons), has been studied in a number of theoretical papers, culminating in the treatment given by Bethe and Heitler in 1934.¹ The radiant energy emitted per centimeter by a particle of mass M carrying a charge ze (e = electronic charge), moving with kinetic energy E through material composed of atoms of atomic number Z , is found to be represented by an expression of the type

$$\frac{z^4 Z^2}{M^2} E f\left(\frac{E}{M}\right), \quad (329)$$

where f denotes a function that is the same for all particles and materials. At high energies, f tends to become constant; the rate of emission of radiant energy is then nearly proportional to the energy of the particle. The loss of energy by radiation decreases rapidly with increase in the mass M of the particle; hence, it is usually important only in the case of electrons. Because of the factor Z^2 , the rate of loss by radiation is much larger in matter composed of heavy atoms, such as lead, than in matter composed of light atoms.

For our purpose it will be convenient to express the rate of loss of energy in terms of an equivalent mean free path for radiation ξ . This is a distance such that, if the loss by radiation were to remain constant, the particle would lose all of its kinetic energy in going a distance ξ ; or, in going a distance dx at initial energy E , the expectation of loss of energy by emission of radiation is $E dx/\xi$. According to Eq. (329), we can write

$$\xi = \frac{M^2}{z^4 Z^2} F\left(\frac{E}{M}\right), \quad (330)$$

where $F = 1/f$.

An actual particle will not radiate all of its kinetic energy in the distance ξ , of course, because, as the energy decreases, the rate of emission decreases. Furthermore, the radiation is emitted in quanta or photons; what has just been said has reference to the mean expectation of the emission of photons, not to a continuous emission of radiation as in classical theory. A photon may contain any fraction

¹ BETHE and HEITLER, *Roy. Soc., Proc.*, vol. 146, p. 83 (1934). Cf. also HEITLER, *loc. cit.*

of the kinetic energy of the particle, from 0 to almost 1. According to the theory, only slightly more energy is radiated, on the average, in the form of photons containing less than half of the kinetic energy than in photons each containing more than half. If we *count* the photons emitted by a group of particles, however, we shall find more of the small ones than of the large ones; the mean number of photons per unit of energy varies a little faster than the reciprocal of $h\nu$.

In Fig. 233, the energy lost per cm. of path, according to the theoretical formulas, is plotted against the logarithm of the energy E of the particle. If y is the ordinate of the curve, the expectation of loss of energy by a particle moving with energy E in going a distance dx is ydx ; or N such particles lose as a group $Nydx$ units of energy. As explained under the figure, some curves refer to the total loss including the loss due to radiation, others to the loss due to ionization and excitation of atoms alone; and, for electrons in air, a curve is shown for the loss by ionization and excitation when those collisions are omitted in which the electron loses more than 10^4 ev in a single collision, the loss so defined corresponding to the "probable" ionization defined in Sec. 235(b). Curves are shown for electrons in lead, and, in standard air, for electrons, for protons, and for mesotrons (Sec. 238) 160 times as heavy as an electron, the radiative energy loss for the latter two kinds of particle being negligible over the range of energies considered.

The curves illustrate strikingly the fact that the radiation loss predominates at high energies (for electrons, above 100 Mev in air or 10 Mev in lead). At energies exceeding the limits for the curves shown, an electron whose kinetic energy is $E \times 10^6$ ev radiates on the average about $30 E$ ev per cm. in standard air, or $1.95 E \times 10^6$ ev per cm. in lead.

By integrating the area under the curve giving the total energy loss, we can find the range of an electron in a given material as a function of its initial kinetic energy, or the total distance that it will travel before coming to rest. Heitler¹ gives the following values of the range in centimeters for an electron moving through standard air or through lead, the initial kinetic energy E being expressed in terms of mc^2 , the electronic rest energy, as a unit:

E	0.1	1	10	100	1,000 mc^2
Air.....	4.4	160	20×10^2	17×10^3	63×10^3
Pb.....	9×10^{-4}	3.7×10^{-2}	0.30	1.25	2.5

¹ HEITLER, *loc. cit.*

(d) *The Growth and Decay of Showers.*—The application of the theoretical results just described to the growth and decay of showers is complicated by the great *straggling* of the elementary processes. Pairs and photons may be produced quickly, or their production may be delayed; and the energy of individual particles or photons may vary over a wide range. For this reason, to follow the course of a shower mathematically presents a difficult problem which has been solved only roughly.¹ We can cite only the results of the analysis. Before doing this, however, it may be instructive to describe a simplified model of the shower process which predicts correctly many of its principal features.

For this purpose, let us assemble rough values of a few of the quantities that are important in shower production, as follows:

	Standard air	Al	Pb
L (for $h\nu = 1,000$ Mev).....	4.5×10^4	13	0.7 cm.
ξ (for radiation).....	3.3×10^4	10	0.5 cm.
E_i (ionization = radiation).....	120	65	9 Mev
R_i	280×10^2	8	0.5 cm.
E_{cp}	25	16	5 Mev

Here L is the mean free path for pair production, and ξ is the corresponding quantity for the emission of radiation by an electron, as just defined. E_i is the energy of an electron at which its mean loss of energy due to inelastic collisions (ionization and excitation) equals that due to the emission of radiation; and R_i is the approximate range of such a particle in the material in question, or the distance it will go before stopping if it starts with energy E_i . E_{cp} is the photon energy at which the probability of loss of energy from the photon by production of a Compton electron equals the probability of loss due to pair production.

To develop a simplified theory, let us now assume, arbitrarily, that each charged particle produced as one member of a pair goes a distance ξ and then emits all of its kinetic energy in 1 or more photons. Since the average frequency at which radiation is actually emitted, as stated above, is about $\nu/2$, where $h\nu = E$, the energy of the particle, we might expect about 2 photons from each particle. The number of photons increases rapidly, however, as the frequency is decreased;

¹ CARLSON and OPPENHEIMER, *Phys. Rev.*, vol. 51, p. 220 (1937); BHABHA and HEITLER, *Roy. Soc., Proc.*, vol. 159, p. 1 (1937); SNYDER, *Phys. Rev.*, vol. 53, p. 960 (1938).

a given amount of energy at a frequency $\nu/4$, *e.g.*, would make 4 photons, as against $\frac{4}{3}$ at frequency $3\nu/4$. A better estimate, therefore, should be 3 photons per particle. Each photon, we will assume, goes a distance exactly equal to L and then turns into two particles, whose energies we suppose to be equal, as they actually are on the average. The net result, on this simplified picture, will then be that in a distance $\xi + L$ one particle is replaced by $3 \times 2 = 6$ particles. This is equivalent to an average rate of multiplication in proportion to $e^{\gamma x}$, where x is the distance covered from the beginning of the shower and $e^{\gamma(\xi+L)} = 6$; since, as appears from the data given above, $L = 1.4 \xi$ roughly, this makes $e^{2.4\gamma\xi} = 6$ and

$$\gamma = \frac{\log 6}{2.4\xi} = \frac{1}{1.34\xi}$$

In an actual shower, of course, photons and particles will coexist. But photons are produced more rapidly than particles, and small photons are produced especially rapidly; hence, we might expect to find about twice as many photons as particles in a shower.

To allow roughly for the effects of ionization, we might now make the simple assumption that the shower stops when the energy of each particle equals E_i , the value at which losses by ionization become equal to losses by radiation. Assuming that particles and photons have equal energies, we should thus obtain for the maximum number of particles in a shower $(\frac{1}{2}) (E_0/E_i)$, E_0 being the initial total energy. The distance covered up to this point would be x_1 , where

$$e^{\gamma x_1} = e^{x_1/(1.34\xi)} = \frac{E_0}{E_i},$$

or $x_1 = 1.34 \xi \log (E_0/E_i)$, E_0/E_i representing the total number of particles and photons. Actually, however, ionization losses and straggling would stop the net growth of the shower somewhat before the point x_1 is reached and, on the other hand, would keep it going with diminishing strength for a considerable distance beyond. We may try to make a first correction for straggling by assuming that at the point x_1 half of the particles have sunk below E_1 in energy and have been stopped. We are thus led to put down as rough estimates of the maximum number N_m of charged particles in the shower and of the distance x_m from the starting-point at which the maximum number occurs

$$N_m = \frac{1}{6} \frac{E_0}{E_i}, \quad x_m = \xi \log \frac{E_0}{E_i}. \quad (331a,b)$$

The corresponding values derived from the mathematical analysis, according to Serber,¹ are well represented, for $E_0/E_i > 20$, by the equations

$$N_m = 0.4 \frac{E_0}{E_i} \left[1 + 1.6 \left(\log \frac{E_0}{E_i} - 1 \right) \right]^{-1/2}, \quad (332a)$$

$$x_m = \xi \left[\log \frac{E_0}{E_i} - \frac{0.8 \log \frac{E_0}{E_i} + 0.5}{\log \frac{E_0}{E_i} - 1} \right]. \quad (332b)$$

A few values calculated from these equations are:

E_0/E_i	25	50	100	250	500	1,000
$\log (E_0/E_i)$	3.22	3.91	4.61	5.52	6.21	6.91
N_m	4.7	8.4	15	35	65	124
x_m/ξ	1.8	2.7	3.4	4.4	5.2	5.9

These values are of the same order of magnitude as the rough estimates given by the easily remembered Eqs. (331a,b).

Additional features predicted by our simplified theory and confirmed by the more exact analysis are the following:

1. The general course of the shower is the same whether it is started by a particle or by a photon.

2. Showers in different materials differ chiefly in the spatial scale of the phenomenon. The maximum occurs in lead at 1 to 3 cm., depending on the energy of the initiating ray, in iron at four times as great a distance, and in standard air at $\frac{2}{3}$ to 2 kilometers. The whole range of the shower will be several times as great as the distance to the maximum.

3. The number of photons in a shower should be nearly double the number of particles.

4. The maximum number of particles should increase linearly with E_0 (actually a little more slowly); and it may attain any magnitude if the initial energy is sufficiently large.

One other interesting feature, predicted by the more exact theory and due especially to the relatively rapid production of small photons, is:

5. The rays composing the shower soon come to be distributed in energy roughly in proportion to $1/E^2$. Such a distribution has been found to agree well with statistical studies in the cloud chamber.

¹ SERBER, *Phys. Rev.*, vol. 54, p. 317 (1938).

The most interesting application of shower theory, perhaps, is to the upper atmosphere of the earth. The number of cosmic-ray particles traveling within 20° of the vertical has been measured at high altitudes, by Pfitzer in latitude 49° , and by Carmichael and Dymond near the north magnetic pole.¹ To make these observations, triple-coincidence arrangements of counters were sent aloft carried by pilot balloons; in the work of Carmichael and Dymond the counts were transmitted to the ground by radio, and the atmospheric pressure was inferred from the observed wave length of transmission. Closely similar curves were obtained at the two locations, rising to a maximum at 8 cm. Hg or about a tenth of the way below the top of the atmosphere, and dropping to half of the maximum value at 2 cm. in the case of Pfitzer's observations. The upper parts of these curves agree well in shape with a theoretical curve drawn to represent the number of particles at various depths in a shower assumed to be initiated at the top of the atmosphere by an electron of 7×10^9 ev energy. Since this energy lies in the range that is inferred for part of the primary cosmic rays from the latitude effect (Sec. 233), the agreement of the theoretical and experimental curves supports the hypothesis that the primary cosmic rays consist in large part of electrons which produce showers in the atmosphere and that the high ionization in the upper part of the atmosphere is mostly due to the electrons and positrons in these showers.

Before leaving the subject of showers, two final remarks concerning the observations should be made.

Occasionally a charged particle appears to eject an atomic electron from an atom with very high energy; thus one particle may be seen to enter a plate of material and two to leave it. If the incident particle is negative, the nature of such an occurrence is unmistakable. If the incident particle is a positron, however, it may be impossible to tell whether the emergent pair of particles arose in the manner described or represents a true shower of two particles, the incident particle having stopped in the plate after emitting a high-energy photon.

Our discussion has been limited, furthermore, to showers of the ordinary type, in which all tracks are "well collimated" or not too far from parallel (an average semiangle of divergence of 10 to 15°). Occasionally showers are observed in the cloud chamber in which the tracks do not seem to be collimated at all, and in such cases heavy tracks ascribed to recoil atoms are often seen. Such showers are

¹ PFOTZER, *Zeits. f. Physik*, vol. 102, pp. 23, 41 (1936); CARMICHAEL and DYMOND, *Nature*, vol. 141, p. 910 (1935).

believed to be due to nuclear disintegrations induced by the cosmic rays in some manner not yet well understood.

238. Mesotrons.—After the discovery of the latitude effect on cosmic-ray ionization, the belief became general that the primary cosmic rays consist, at least in large part, of (negative) electrons. It was also believed that the cosmic-ray tracks seen in cloud chambers are mostly made either by these primary electrons or by secondary electrons or positrons.

By the end of 1934, however, several difficulties with this interpretation had emerged.

Let us consider what energy an electron would have to have in order to penetrate the whole depth of the atmosphere. If we multiply an ionization loss of about 2,500 ev per cm. by the thickness of the atmosphere in terms of standard air, or 8×10^5 cm., we obtain 2×10^9 ev for the total loss. This is less than the energies with which most of the primary rays enter the atmosphere, as inferred from the latitude effect (up to 10^{10} ev or more—see Sec. 233).

So far, therefore, the assumption that the primary cosmic rays are electrons meets with no difficulty. The picture is changed completely, however, when losses by radiation are included. The total loss becomes then, roughly, $2,500 + 3 \times 10^{-5}E$ ev cm.⁻¹, E denoting the energy of the electron and the second term representing the mean loss of energy due to radiation according to the Bethe-Heitler formula [Sec. 237(c)]. Thus in a distance dx the mean loss is

$$dE = -(a + bE)dx,$$

where $a = 2,500$, $b = 3 \times 10^{-5}$; whence, upon integrating,

$$E = \frac{a}{b} \left[\left(1 + \frac{bE_0}{a} \right) e^{-bx} - 1 \right],$$

E_0 being the initial energy in ev at $x = 0$. Thus E becomes zero, and the electron stops moving, when

$$1 + \frac{bE_0}{a} = e^{bx}, \quad x = \frac{1}{b} \log \left(1 + \frac{bE_0}{a} \right)$$

or

$$x = 3.3 \times 10^4 \log(1 + 1.2 \times 10^{-8} E_0) \text{ cm.}$$

To make $x = 8 \times 10^5$ cm., so that the particle can just reach the earth, we should have to have

$$\log(1 + 1.2 \times 10^{-8} E_0) = \frac{80}{3.3} = 24, \quad E_0 = 2.2 \times 10^{18} \text{ ev.}$$

If E_0 were only 10^{11} ev, which is already far above the upper limit of 6×10^9 for the latitude effect, the range in standard air would be only 2.4 km. Thus electrons or positrons with energy of ordinary cosmic-ray magnitude could not possibly get through the whole thickness of the atmosphere.

Similar difficulties were revealed by laboratory experiments. As we have seen in Sec. 234, Rossi detected charged particles capable of passing through a meter of lead, which is equivalent, as regards theoretical energy loss, to more than the whole atmosphere. Furthermore, as we have pointed out before, with a cloud chamber in a magnetic field, it is possible to determine directly the loss of energy by a particle in passing through a layer of matter; it is only necessary to measure the difference in the curvatures of the tracks above and below the layer of matter, from which the change in energy can be calculated [with a possible error of a few percent in case the mass of the particle is uncertain—see [Sec. 235(a)]]. Extensive measurements of this sort were begun by Anderson in 1932, at Millikan's suggestion, and were continued by Anderson and Neddermeyer¹ and others.² In such observations it was found very difficult to secure consistent results. In some cases, energy losses were observed that agreed well with the Bethe-Heitler theory of radiation, which had been brought out in 1934. In other cases, especially for high-energy particles, the loss was certainly much less. For a few years, it was believed that all of the known facts taken together pointed toward a breakdown of the wave-mechanical theory of radiation for electron energies above 10^8 electron-volts. The development of the theory of showers was probably delayed for a time by this erroneous conclusion.

As the experimental work progressed, however, it gradually became clear that all cosmic-ray particles are not alike. In 1936, Anderson and Neddermeyer reported observations on the amount of energy lost by cosmic-ray particles in passing through a lead plate 3.5 mm. thick placed across a cloud chamber. They found that, although the individual variations were enormous (presumably due to straggling in the emission of photons), the *average* loss by *particles forming part of showers* was in approximate agreement with the theory.³ For a further test, Neddermeyer and Anderson replaced the lead plate by one of platinum 1 cm. thick, equivalent to nearly 2 cm. of lead.⁴ It was found that shower particles of energy below 500 Mev

¹ ANDERSON, *Phys. Rev.*, vol. 41, p. 405 (1932), and later issues of the *Phys. Rev.*

² For references see *Rev. Modern Physics*, vol. 10, p. 174 (1938).

³ ANDERSON and NEDDERMEYER, *Phys. Rev.*, vol. 50, p. 263 (1936).

⁴ NEDDERMEYER and ANDERSON, *Phys. Rev.*, vol. 51, p. 884 (1937).

could get through even such a thick plate in small numbers, but they emerged with almost a total loss of their energy. This can be understood if these particles are electrons or positrons and are able to traverse the plate only because their radiative energy losses happen to be abnormally low. The particles occurring singly, on the other hand, lost, as a rule, less than half of their energy in traversing the platinum.

The data in question and some later ones are shown in Fig. 234. The abscissa represents the energy with which a particle entered the

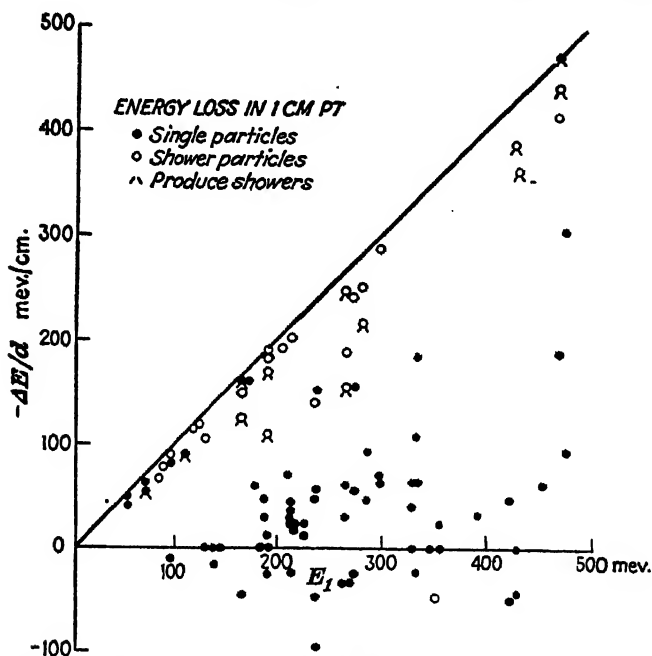


FIG. 234.—Energy loss in 1 cm. of platinum as function of incident energy. (Energy calculated on assumption of electronic mass.) (Anderson and Neddermeyer.)

cloud chamber, the ordinate, its loss of energy in passing through 1 cm. of platinum. Circles refer to shower particles, solid dots to particles that occurred alone. All points referring to particles that traverse the platinum must necessarily lie below the line drawn in the diagram, since this line represents total loss of energy in the plate. The plot furnishes clear evidence for the conclusion that at least two different kinds of cosmic-ray particles occur in cloud chambers, characterized by wide differences in penetrating power. The occurrence of cases of an *apparent gain* of energy, however, well illustrates the difficulty of making accurate measurements of this type. Highly

penetrating particles, with energies up to 10^9 ev, have been reported also by Crussard and Leprince-Ringuet.¹

All of the facts can be brought into harmony with the wave-mechanical theory of radiation if we assume that the exceptionally *penetrating particles* have a *much smaller ratio of charge to mass* than has an electron. As we have seen, a satisfactory theory of *showers* can be developed, using wave mechanics, on the assumption that showers consist of electrons, positrons, and photons. Furthermore, it is possible to interpret the soft or nonpenetrating component of the cosmic rays, defined as that part of the rays which is absorbed by about 10 cm. of lead [Sec. 236(c)], as consisting of shower rays that have been produced in solid objects or in the atmosphere above the point of observation. The *penetrating component* of the cosmic rays is then presumably to be identified with the penetrating particles observed in cloud chambers. What is the nature of these particles?

Can they be protons? A proton would suffer no appreciable loss by radiation in passing through matter, losing energy only by ionization, and so would be much more penetrating than an electron. The following objections can be raised, however, against the assumption that the particles in question are protons:

1. Since more than a third of them carry a negative charge, it would be necessary to admit the existence of negative protons, for whose existence there is no other evidence.

2. The tracks of the penetrating particles are scarcely distinguishable from electron tracks, whereas a proton of energy below 2×10^8 ev should make a track many times heavier. If the penetrating particles are protons, they ought often to be seen close to the end of their range, moving slowly enough to cause dense tracks. Careful search by several investigators has failed to reveal an appreciable number of such tracks in cloud chambers.²

3. The distribution in energy of the high-speed electrons ejected from atoms by penetrating particles does not agree well with that obtained from theory on the assumption that these particles are protons, but would be consistent with a mass several times smaller.³

In view of these considerations, it is thought that the penetrating cosmic-ray particles are probably a new type of particle, with a mass intermediate between the masses of the electron and the proton. From Eq. (329), it is evident that the mass would only have to be 10 to 20 times the electronic mass in order to reduce the radiation loss

¹ CRUSSARD and LEPRINCE-RINGUET, *J. de Physique*, vol. 8, p. 213 (1937)

² Cf. *Rev. Modern Physics*, vol. 10, p. 180 (1938).

³ *Ibid.*, vol. 11, p. 203 (1939).

to small proportions and so make the particles highly penetrating. Various names have been proposed for these new particles (dynaton, yukon, barytron, heavy electron); the drift is toward the name *mesotron* or, perhaps, *meson*.¹

The mass of the *mesotron* has been estimated mostly from its ionizing power in the gas of the cloud chamber. As stated in Sec. 235(b), a particle carrying the same charge as an electron but k times as heavy should ionize almost like an electron having an energy $1/k$ times as great. In 1937, Street and Stevenson² obtained a track in a cloud chamber under 11 cm. of lead with $H\rho = 9.6 \times 10^4$ oersted-cm. and a specific ionization about 6 times as great as in normal thin tracks. If the particle entered from above, its sign was negative. The heavy ionization would be explained if this particle had a mass 130 ± 30 times the electronic mass. It certainly was not a proton, for, if it had been, its energy would have been only 4.5×10^6 ev [by Eq. (328b)]; a proton of this energy is known to have a range of only 1 cm. in standard air, whereas the track in question was clearly visible over a length of 7 cm. A thousand counter-controlled expansions were necessary to obtain this single useful track.

Other estimates of the mass of the *mesotron* have ranged from 100 to 400 times the electronic mass; a favorite guess is about 180. Perhaps the mass is not a fixed quantity. Only further experiments can answer this question.

239. Origin and Fate of Mesotrons.—If mesotrons exist, where do they come from, and what becomes of them? There is at present no evidence that they play any role in ordinary nuclear transformations or in any other laboratory phenomena.

There is some evidence, furnished by airplane observations at 25,000 ft., that mesotrons may be produced by photons.³ An indirect argument of another sort, on the other hand, indicates that they may be produced principally by *primary protons*.

At low latitudes, more mesotrons are observed to come from directions west of the zenith than from directions east of it. Now, as was explained in Sec. 232, the theory of the motion of charged particles through the earth's magnetic field indicates that such particles, if their energy lies in a certain range, should arrive at the earth predominantly from the west or from the east according as they carry positive or negative electrical charges. Observation shows that about

¹ *Mesotron* is apparently formed from the Greek root "meso," middle, and a meaningless fragment of the word, electron.

² STREET and STEVENSON, *Phys. Rev.*, vol. 52, p. 1003 (1937).

³ SCHEIN and WILSON, *Rev. Modern Physics*, vol. 11, p. 292 (1939).

57 percent (or 73 percent?) of the mesotrons are positively charged. Such a moderate predominance of positives, however, does not seem to be sufficient to account for the observed "east-west asymmetry" in their directions of arrival if we assume that the mesotrons are themselves the primary particles. From an analysis of all of the data, Johnson concludes that the part of the cosmic-ray ionization which varies with latitude must be *entirely due*, directly or indirectly, to *positive* primary particles.¹ If this is true, the mesotrons must represent secondary particles of some sort produced in the earth's atmosphere.

At this point speculative theory steps in with the suggestion that protons colliding with other protons or with neutrons in the nuclei of atoms may be able to produce mesotrons in a manner rather analogous to the production of photons by electrons in colliding with nuclei (e.g., in the emission of continuous X-rays).² Thus we arrive at the hypothesis that the mesotrons are produced by primary protons entering the earth's atmosphere from the outside.

This conclusion is supported by recent observations by Herzog, who succeeded in taking cloud-chamber photographs on an airplane at 29,000 ft. and found many heavy tracks which he ascribed to protons.³

Unlike the mesotrons, showers do not exhibit an east-west asymmetry of any magnitude, at least not at high altitudes.⁴ Taken in conjunction with the shower theory, this fact suggests that the primary particles which cause most of the showers at high altitudes consist of electrons and positrons in about equal proportions.

As to the ultimate fate of the mesotrons, the theoretical speculations just mentioned indicate that they should spontaneously turn into electrons or positrons, according to the sign of their charge; part of their energy, including the excess of their rest energy, will then be either emitted in radiation or carried off by a neutrino (Sec. 212). There is as yet (1940) no satisfactory experimental evidence that mesotrons do become electrons or positrons. Several phenomena are known, however, which are most easily explained on the assumption that for some reason or other the mesotrons possess a limited mean life in the atmosphere.

The slight *decrease* of the cosmic-ray ionization with *rise of temperature*, mentioned above (Sec. 233), would be accounted for if the

¹ JOHNSON, *Rev. Modern Physics*, vol. 10, p. 232 (1938).

² For references see *Rev. Modern Physics*, vol. 10, p. 176 (1938).

³ HERZOG, *Phys. Rev.*, vol. 59, p. 117 (1941).

⁴ Cf. *Rev. Modern Physics*, vol. 10, p. 223 (1938).

mesotrons decay spontaneously. For, a higher temperature, at given barometric pressure, means that the atmosphere extends farther upward from the earth, so that the mesotrons produced in it have farther to go and more of them perish before reaching the earth.¹

Data of another sort pointing to the existence of mesotron decay have been obtained by Rossi, Hilberry, and Hoag.² Using a vertical counter telescope, with 12.7 cm. of lead between the counters in order to stop the soft component of the cosmic-ray particles, these experimenters compared the mesotron intensities at four different elevations near latitude 50°, viz., at Chicago (180 meters) and stations at 1,616, 3,240, and 4,300 meters in Colorado. The observed increase of mesotron intensity with altitude was ascribed to a combination of absorption and disintegration of the mesotrons in the intervening vertical air mass. At each station except Chicago, the absorption in a thickness of graphite equal to 87 grams cm.⁻² was also determined. Carbon has the same ratio of electrons to mass as have nitrogen and oxygen; therefore, the true absorption of mesotrons in the two mediums should be about the same. Nevertheless, only about half as many mesotrons were found to be stopped by the graphite as by an equal mass of atmospheric air. The additional apparent absorption in the air may be due to spontaneous disintegration of the mesotrons during the longer time it takes them to traverse the air.

On this assumption, it was calculated from the data that a mesotron of cosmic-ray energy would travel in a vacuum an average distance of 9.5 km., occupying a time of 3.2×10^{-5} sec., before disintegrating. For a mesotron at rest, the mean life would then be about 2×10^{-6} sec., the longer time for the moving one being due to the relativistic effect on time scales (Sec. 64). The temperature effect referred to above would be explained by a mean life at rest of about 3.4×10^{-6} sec.

It may be remarked that the data just mentioned gave for the apparent mass absorption coefficient of the mesotrons in air, near sea level (70 cm. Hg), $\mu/\rho = 0.8 \times 10^{-3}$, increasing to 2.4×10^{-2} at 48 cm. Hg. In carbon, the data gave $\mu/\rho = 0.4 \times 10^{-3}$ near sea level. This latter figure is in rough agreement with the value of 0.55×10^{-3} found by Rossi for the penetrating component of the cosmic rays in lead near sea level (Sec. 234). The two measurements are probably not strictly comparable because of differences in the loss of intensity due to scattering in the two cases.

¹ BLACKETT, *Phys. Rev.*, vol. 54, p. 973 (1938).

² ROSSI, HILBERRY, and HOAG, *Phys. Rev.*, vol. 57, p. 461 (1940).

The study of mesotrons is obviously in its infancy. The student should consult the literature of the subject from 1940 on for further developments.

240. Conclusion.—In this chapter, as proposed at the beginning, we have touched only upon certain topics in the field of cosmic rays concerning which definite results appear to have been achieved. Many other topics might have been discussed. A few of these may be mentioned in conclusion.

The student may have wondered whether the shower particles which have been assumed to constitute the soft or easily absorbable component of the cosmic-ray particles at the earth's surface have been produced by mesotrons in the neighborhood of the point of observation or constitute the remnants of showers produced by incoming electrons in the upper atmosphere. It does not seem possible that particles from the latter showers could get through to the earth in sufficient numbers; but this point is still (1940) obscure.

In this connection, mention may be made of the interesting observations of Auger and others on *very broad showers*.¹ Coincidences were obtained with two or three counters placed as much as 10 meters apart. Studies of the results suggested the interpretation that the particles responsible for such coincidences belong to showers containing up to a million particles, each produced by a single primary particle with an enormous energy of the order of 10^{15} ev. About 60 such showers seemed to arrive per hour at sea level, but 600 per hour on the Jungfrauoch (3,500 meters).

Certain *unusually narrow and penetrating showers*, on the other hand, have been interpreted by Bothe as *showers of mesotrons*.²

Neutrons, also, have been found by various investigators in the atmosphere. In observations by Korff and others,³ they were found to increase from about 6×10^{-9} slow neutrons per cm.² at sea level to nearly 4×10^{-6} per cm.² at 1 meter of water below the top of the atmosphere (altitude, 10 miles). The neutrons were detected with a counter filled with BF₃ at 0.1 atmosphere pressure [cf. Sec. 223(b)]; the counter and a barometer were carried aloft by pilot balloons and their readings were automatically transmitted to a ground station by short-wave radio. The neutrons may be entering the earth's atmosphere from the outside, or they may be produced in it by processes of nuclear disintegration.

¹ AUGER, *Comptes Rendus*, vol. 206, p. 1721 (1938); vol. 207, p. 228 (1938); *J. de Physique et le Radium*, vol. 10, p. 39 (1939); *Rev. Modern Physics*, vol. 11, p. 288.

² BOTHE, *Rev. Modern Physics*, vol. 11, p. 282 (1939).

³ Cf. KORFF, *Rev. Modern Physics*, vol. 11, p. 211 (1939).

Small variations in cosmic-ray intensity at a given station have often been detected, but their significance is uncertain.¹ There appears to be a diurnal variation with a maximum at noon and a minimum at midnight, the extreme range being roughly 0.4 percent. This variation may be caused by cosmic rays coming directly from the sun, or it may result from the deflecting effects of the sun's magnetic field upon particles approaching from outside of the solar system.

Such variations are obviously of great importance in connection with the problem of the *origin of the primary cosmic rays*. Where do they come from, and how do they acquire their enormous energies? No convincing theory has been proposed. In so far as they consist of charged particles, they have presumably been accelerated in electric fields of enormous extent, existing in stars or throughout space. Possibly they are produced in the new stars of explosive type called "supernovae."

Many questions still unmentioned will probably have occurred to the student. In the literature he will find answers, or attempts at answers, to some of these questions. Others will doubtless form the subject of future investigations. The last section of this book might better be called not a "conclusion" but an "introduction to future research." Further surprises may be in store for the physicist in the fascinating field of cosmic rays.

¹ See *Rev. Modern Physics*, vol. 11, pp. 153-190 (1939).

APPENDIX I

ISOTOPIC CONSTITUTION OF THE ELEMENTS

The data in this Appendix have been taken from the sources listed below, relative abundances of isotopes and chemical atomic weights being taken from the first paper listed.

HARN, FLÜGGE, and MATTAUCH, *Phys. Zeits.*, vol. 41, p. 1 (1940); BARKAS, *Phys. Rev.* 55, p. 691 (1939); DEMPSTER, *Phys. Rev.*, vol. 53, pp. 74, 869 (1938); ASTON, *Roy. Soc., Proc.*, vol. 163, p. 391 (1937); POLLARD, *Phys. Rev.*, vol. 57, p. 1186 (1940).

In computing the chemical atomic weight from the isotopic weights and the data on relative abundance, the factor $16/16.0044 = 1/1.000275$ was used to convert from a.m.u. to chemical units.

Element Z	Mass number A of isotopes	Atomic mass, a.m.u., (O ¹⁶ = 16)	Packing fraction × 10 ⁴	Relative abundance, percent	Chemical atomic weight (O = 16)	
					Computed from relative abundance	Observed
H 1	1	1.00812	81.2	99.98	1.0080	1.0081
	2	2.01472	73.6	0.02		
He 2	4	4.00388	9.8	100	4.0027	4.003
Li 3	6	6.0169	28.2	7.9	6.937	6.940
	7	7.0180	25.7	92.1		
Be 4	9	9.0150	10.7	100	9.013	9.02
B 5	10	10.0160	16.0	20	10.81	10.82
	11	11.0129	11.7	80		
C 6	12	12.0040	3.3	98.9	12.012	12.010
	13	13.0077	5.9	1.1		
N 7	14	14.00750	5.4	99.62	14.007	14.008
	15	15.0049	3.3	0.38		
O 8	16	16.0000	0	99.76	16.0000	16.0000
	17	17.0045	2.7	0.04		
	18	18.005	3	0.20		
F 9	19	19.0045	2.4	100	18.9993	19.00
Ne 10	20	19.9988	-0.6	99.00	20.101	20.183
	21	20.9997	-0.1	0.27		
	22	21.9986	-0.6	9.73		
Na 11	23	22.9961	-1.7	100	22.990	22.997
Mg 12	24	23.0024	-3.2	77.4	24.323	24.32
	25	24.9938	-2.5	11.5		
	26	25.9898	-3.9	11.1		
Al 13	27	26.990	-3.7	100	26.983	26.97
Si 14	28	27.987	-4.6	89.6	28.125	28.06
	29	28.987	-4.5	0.2		
	30	29.983	-5.7	4.2		
P 15	31	30.984	-5.2	100	30.977	30.98
S 16	32	31.9823	-5.5	95.0	32.065	32.06
	33	32.983	-5.2	0.74		
	34	33.978	-6.5	4.2		
	36	35.978	-6.1	0.016		
Cl 17	35	34.980	-5.7	75.4	35.462	35.457
A 18	37	36.978	-5.9	24.6		
	36	35.978	-6.1	0.31		
K 19	38	37.974	-6.8	0.06	39.026	39.944
	40	39.9750	-6.2	99.63		
	39	38.975	-6.4	93.44	39.007	39.006
	40*	39.975	-6.5	0.012		
Ca 20	41	40.974	-6.4	6.55	40.08	40.08
	40	39.974	-6.5	96.96		
	42	41.971	-6.9	0.64		
	43	42.972	-6.5	0.15		
	44	2.07		
	46	0.003		
	48	0.185		

Element <i>Z</i>	Mass number <i>A</i> of isotopes	Atomic mass, a.m.u. ($O^{16} = 16$)	Packing fraction $\times 10^4$	Relative abundance, percent	Chemical atomic weight ($O = 16$)	
					Computed from relative abundance	Observed
Sc 21 Ti 22	45	44.969	-6.9	100	44.96 47.88	45.10 47.90
	46	45.968	-7.0	7.95		
	47	7.75		
	48	47.965	-7.2	73.45		
	49	48.966	-7.2	5.51		
	50	49.963	-7.4	5.34		
V 23 Cr 24	51	50.960	-7.9	100	50.95 52.00	50.95 52.01
	50	4.49		
	52	51.959	-7.9	83.78		
	53	52.957	-8.1	9.43		
Mn 25 Fe 26	54	2.30	54.94 55.85	54.93 55.84
	55	54.965	-8.1	100		
	54	53.960	-7.4	6.0		
	56	55.961	-7.0	91.6		
	57	56.961	-6.9	2.1		
Co 27 Ni 28	58	0.28	58.94 58.69
	58	68.0		
	60	59.959	-6.8	27.2		
	61	0.1		
	62	3.8		
	64	0.9		
Cu 29	63	62.956	} mean, {	68	63.58	63.57
	65	64.955		32		
Zn 30	64	63.956	-6.9	50.9	65.38	65.38
	66	65.952	-7.4	27.3		
	67	3.9		
	68	67.956	-6.5	17.4		
	70	69.954	-6.5	0.5		
Ga 31	69	68.955	-6.5	61.5	69.70	69.72
	71	70.953	-6.6	38.5		
Ge 32	70	21.2	72.60
	72	27.3		
	73	7.9		
	74	37.1		
	76	6.5		
As 33 Se 34	75	100	74.91 78.96
	74	0.9		
	76	9.5		
	77	8.3		
	78	24.0		
	80	48.0		
Br 35	82	9.3	79.916	79.916
	79	50.6		
	81	49.4		
		
Kr 36	78	77.945	-7.0	0.35	83.82	83.7
	80	2.01		
	82	81.939	-7.5	11.52		
	83	11.52		
	84	83.938	-7.3	57.13		
	86	85.939	-7.1	17.47		
Rb 37	85	72.8	85.48
	87*	27.2		
Sr 38	84	0.56	87.63
	86	9.86		
	87	7.02		
	88	82.56		
Y 39 Zr 40	89	100	88.92 91.22
	90	48.0		
	91	11.5		
	92	22		
	94	17		
	96	1.5		
		

Element Z	Mass number A of isotopes	Atomic mass, a.m.u. (O ¹⁶ = 16)	Packing fraction × 10 ⁴	Relative abundance, percent	Chemical atomic weight (O = 16)	
					Computed from relative abundance	Observed
Cb 41	93	100	92.91
Mo 42	92	15.5	95.95
	94	8.7		
	95	94.945	-5.8	16.3		
	96	95.946	-5.6	16.8		
	97	96.945	-5.8	8.7		
	98	97.944	-5.7	25.4		
	100	8.6		
43						
Ru 44	96	95.946	-5.7	(5)	101.7
	98	—?		
	99	98.944	-5.7	(12)		
	100	(14)		
	101	(22)		
	102	(30)		
	104	(17)		
Rh 45	103	100	102.91
Pd 46	1028	106.55	106.7
	104	103.946	-5.2	9.3		
	105	104.945	-5.2	22.6		
	106	105.945	-5.2	27.2		
	108	107.943	-5.2	26.8		
	110	109.942	-5.2	13.5		
Ag 47	107	106.948	-4.8	52.5	107.87	107.880
	109	108.947	-4.8	47.5		
Cd 48	106	1.4	112.41
	108	1.0		
	110	12.8		
	111	13.0		
	112	24.2		
	113	12.3		
	114	28.0		
	116	7.3		
In 49	113	4.5		114.76
	115	95.5		
Sn 50	112	1.1	118.70	118.70
	114	0.8		
	115	0.4		
	116	115.942	-5.0	15.5		
	117	9.1		
	118	117.939	-5.1	22.5		
	119	118.938	-5.2	9.8		
	120	119.94	-6	28.5		
	122	121.944	-4.6	5.5		
	124	123.943	-4.6	6.8		
Sb 51	121	56	121.76
	123	44		
Te 52	122	2.9	127.61
	123	1.6		
	124	4.5		
	125	6.0		
	126	19.0		
	128	32.8		
	130	33.1		
	127	100		
I 53	124	0.004	126.92
Xe 54	126	0.088		
	128	1.91		
	129	128.946	-4.2	20.23		
	130	4.06		
	131	21.18		
	132	131.946	-4.4	26.98		
	134	10.55		
	136	8.95		
					(131.3)	131.3

Element <i>Z</i>	Mass number <i>A</i> of isotopes	Atomic mass, a.m.u., ($O^{16} = 16$)	Packing fraction $\times 10^4$	Relative abundance, percent	Chemical atomic weight ($O = 16$)	
					Computed from relative abundance	Observed
Cs 55	133	100	132.91
Ba 56	130	0.101	137.36
	132	0.097		
	134	2.42		
	135	6.6		
	136	7.8		
	137	11.3		
	138	71.7		
La 57	139	138.955	-3.2	100	138.92	138.92
Ce 58	136	<1	140.13
	138	<1		
	140	89		
	142	11		
Pr 59	141	100	140.92
Nd 60	142	25.95	144.27
	143	13.0		
	144	22.6		
	145	9.2		
	146	145.980	-2.8	16.5		
	148	147.961	-2.7	6.8		
	150	149.987	-2.2	5.95		
		
II 61
Sm 62	144	3	150.43
	147	17		
	148*	14		
	149	15		
	150	5		
	152	26		
	154	20		
Eu 63	151	49.1	152.0
	153	50.9		
		
		
		
Gd 64	152	0.2	156.9
	154	1.5		
	155	154.977	-1.5	21		
	156	155.976	-1.5	22		
	157	156.976	-1.5	17		
	158	22		
	160	16		
		
Tb 65	159	100	159.2
Dy 66	158	0.1	162.46
	160	1.5		
	161	22		
	162	24		
	163	24		
	164	28		
		
Ho 67	165	100	164.935
Er 68	162	0.25	167.2
	164	2.0		
	166	(35)		
	167	(24)		
	168	(29)		
	170	(10)		
		
Tm 69	169	100	169.4
Yb 70	168	0.06	173.04
	170	2		
	171	(9)		
	172	(23)		
	173	(17)		
	174	(37)		
	176	(12)		
Lu 71	175	97.5	174.99
	176*	2.5		

Element <i>Z</i>	Mass number <i>A</i> of isotopes	Atomic mass, a.m.u. ($O^{16} = 16$)	Packing fraction $\times 10^4$	Relative abundance, percent	Chemical atomic weight ($O = 16$)	
					Computed from relative abundance	Observed
Hf 72	174	(0.3)		
	176	(5)		
	177	(19)		
	178	(28)		
	179	(18)		
	180	(30)		178.6
Ta 73	181	100	180.88
	180	0.2±		
	182	22.6		
	183	17.3		
	184	30.1		183.92
	186	29.8		
Re 75	185	38.2	186.31
	187	61.8		
Os 76	184	0.018		
	186	1.59		
	187	1.64		
	188	13.3		
	189	16.2		
	190	190.038	+2.0	26.4		190.2
	192	192.038	2.0	40.9		
Ir 77	191	191.040	2.1	38.5	} 192.22	193.1
	193	193.041	2.1	61.5		
Pt 78	192	0.8	} 195.14	195.23
	194	194.040	2.0	30.2		
	195	195.040	2.0	35.3		
	196	196.039	2.0	26.6		
	198	198.044	2.2	7.2		
Au 79	197	197.039	2.0	100	196.99	197.2
Hg 80	196	0.15		
	198	10.12		
	199	17.04		
	200	200.028	1.4	23.25		200.61
	201	13.18		
	202	29.54		
	204	6.72		
Tl 81	203	203.057	2.8	29.1	} 204.42	204.39
	205	205.057	2.8	70.9		
Pb 82	204	204.058	2.8	1.5		
	206	23.6		207.21
	207	22.6		
	208	208.057	2.7	52.3		
Bi 83	209	209.055	2.6	100	209.00	209.00
Po 84
Rn 86	222*	222
— 87
Ra 88	223*	226.05
Ac 89	224*
Th 90	226*
Pa 91	227*
U 92	232*	232.12	5.2	100	232.06	232.12
	231*
	234*	0.006	} 238.06	238.07
	235*	0.720		
	238*	238.14	5.6	99.274		

* Radioactive.

APPENDIX II

THE PERIODIC TABLE

Periods	I	II	III	IV	V	VI	VII	VIII	
I	1 H 1, 2								(a) 40, 44, 42, 48, 43, 46
II	3 Li 7, 6	4 Be 9	5 B 11, 10	6 C 12, 13	7 N 14, 15	8 O 16, 18, 17	9 F 19	10 Ne 20, 22, 21	(b) 80, 78, 76, 82, 77, 74
III	11 Na 23	12 Mg 24, 25, 26	13 Al 27	14 Si 28, 29, 30	15 P 31	16 S 32, 34, 33, 36	17 Cl 35, 37	18 Ar 40, 36, 38	(c) 84, 86, 82, 88, 80, 78
IV	19 K 39, 41, 40	20 Ca (a)	21 Sc 45	22 Ti 48, 46, 50	23 V 51	24 Cr 52, 53, 50, 54	25 Mn 55	26 Fe 56, 54, 57, 58	(d) 98, 96, 95, 92, 94, 97, 100
V	29 Cu 63, 65	30 Zn 68, 67, 70	31 Ga 69, 71	32 Ge 70, 73, 76	33 As 75	34 Se (b)	35 Br 79, 81	36 Kr (c)	(e) 114, 112, 111, 110, 113, 116, 106, 108
	37 Rb 85, 87	38 Sr 86, 87, 84	39 Y 89	40 Zr 92, 94, 91, 96	41 Nb 93	42 Mo (d)	43 Tc (e)	44 Ru 102, 101, 104, 100, 99, 96	(f) 120, 118, 116, 119, 117, 124, 122, 112, 114, 116
	47 Ag 107, 109	48 Cd (c)	49 In 115, 113	50 Sn (f)	51 Sb 121, 123	52 Te (g)	53 I 127	54 Xe (h)	(g) 130, 128, 126, 125, 124, 122, 123
VI	55 Cs 133	56 Ba (i)	57-71 La * 3-17	58 Ce (j)	59 Pr 181	60 Nd 182, 183, 180	61 Pm 187, 185	62 Sm 192, 190	(h) 132, 129, 131, 134, 136, 130, 128, 124, 126
VII	79 Au 197	80 Hg (k)	81 Tl 203, 205	82 Pb 206, 207, 204	83 Bi 209	84 Po 210	85 At 215	86 Rn 222	(i) 138, 137, 136, 135, 134, 130, 132
	87 Rb (223, 224, 226)	88 Sr (223, 224, 226)	89 Y 89	90 Zr 90	91 Nb 91	92 Mo 92	93 Tc 93	94 Ru 94	(j) 180, 178, 177, 179, 176, 174
									(k) 202, 200, 199, 201, 198, 204, 196

Explanation.—The Roman numerals in the top row designate the "columns" of the periodic table. The Roman numerals at the left give the several "periods" according to Bohr's arrangement (Fig. 93), and the italic numerals in the upper right-hand corners of the several spaces give the ordinal numbers of the elements in these periods. Atomic numbers are indicated by bold-faced type in the upper left-hand corners. Mass numbers of the constituent isotopes are given by the numbers at the bottoms of the spaces, in the order of decreasing abundance (except for Ra).

Element	Atomic Weight	Electron Configuration															
		1s				2s				3s				4s			
		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Ce 58		48 electrons silver core															
Pr 59		48 electrons silver core															
Nd 60		48 electrons silver core															
Pm 61		48 electrons silver core															
Sm 62		48 electrons silver core															
Eu 63		48 electrons silver core															
Gd 64		48 electrons silver core															
Tb 65		48 electrons silver core															
Dy 66		48 electrons silver core															
Ho 67		48 electrons silver core															
Er 68		48 electrons silver core															
Tm 69		48 electrons silver core															
Yb 70		48 electrons silver core															
La 71		48 electrons silver core															
Hf 72		68 electrons hafnium core															
Ta 73		68 electrons hafnium core															
W 74		68 electrons hafnium core															
Re 75		68 electrons hafnium core															
Os 76		68 electrons hafnium core															
Ir 77		68 electrons hafnium core															
Pt 78	8.9	68 electrons hafnium core															
Au 79	9.22	68 electrons hafnium core															
Hg 80	10.419	68 electrons hafnium core															
Tl 81	6.008	68 electrons hafnium core															
Pb 82	7.404	68 electrons hafnium core															
Bi 83	8.0	68 electrons hafnium core															
Po 84		68 electrons hafnium core															
At 85		68 electrons hafnium core															
Rn 86	10.780	68 electrons hafnium core															
Fr 87		86 electrons radon core															
Ra 88	10.2	86 electrons radon core															
Ac 89		86 electrons radon core															
Th 90		86 electrons radon core															
Pa 91		86 electrons radon core															
U 92		86 electrons radon core															

Note.—The ionization potentials are mostly calculated, using modern values of the constants [Eq. (147c) in Sec. 108], from data in Bacher and Goudsmit, "Atomic Energy States—As Derived from the Analyses of Optical Spectra," 1932.

SOME USEFUL CONSTANTS AND RELATIONS

(Except in a_0 , mc^2 , M_1c^2 , kT at 15° , the last figure given is uncertain)

c	velocity of light ¹	2.9978×10^{10} cm. sec. ⁻¹
F	the faraday ¹	9,648.9 e.m.u. mole ⁻¹
e	electronic charge ²	4.803×10^{-10} e.s.u.
m	electronic mass (Sec. 45).....	0.9107×10^{-27} gram
e/m	ratio of electronic charge to mass ³	1.7591×10^7 e.m.u. gram ⁻¹
M_{w-1}	mass of atom of atomic weight 1 (Sec. 206)	1.6604×10^{-24} gram
M_1	atomic mass unit (Sec. 206).....	1.6599×10^{-24} gram
h	Planck's constant ⁴	6.610×10^{-27} erg sec.
R	universal gas constant.....	$\left\{ \begin{array}{l} 8.315 \times 10^7 \text{ ergs deg.}^{-1} \text{ mole}^{-1} \\ 82.06 \text{ cm.}^3 \text{ atm. deg.}^{-1} \text{ mole}^{-1} \end{array} \right.$
N_0	Avogadro's number (Sec. 85).....	6.023×10^{23} mole ⁻¹
k	Boltzmann's constant (Sec. 85).....	1.381×10^{-16} erg deg. ⁻¹
σ	Stefan's constant ¹	5.735×10^{-5} erg cm. ⁻² sec. ⁻¹ deg. ⁻⁴
R_∞	Rydberg's wave number ¹	1.0973742×10^5 cm. ⁻¹
M_H/m	ratio of masses of hydrogen atom and electron (Sec. 106).....	1,837
$a_0 = \frac{h^2}{4\pi^2me^2}$	radius of smallest Bohr orbit in hydrogen (Sec. 103).....	5.27×10^{-9} cm.
$\alpha = \frac{2\pi e^2}{ch}$	"fine-structure constant" (Sec. 148)...	0.007314 = 1/136.7
1 electron-volt (ev, Sec. 105)		$= 1.601 \times 10^{-12}$ erg
M_1c^2	rest energy of 1 atomic mass unit (Sec. 214)	932×10^6 electron-volts
mc^2	rest energy of the electron (Sec. 214)	0.511×10^6 electron-volts
kT at 15°C. (T = absolute temperature = 288.1°K.)		0.0248 electron-volts
If $h\nu = eV_v/300$ (V_v = potential difference in volts or energy in electron-volts),		$\lambda = c/\nu$, $\lambda_A = \lambda \times 10^8$ = wave length in angstroms, $\tilde{\nu} = 1/\lambda$ or waves per cm.,

$$\lambda_A = \frac{12,378}{\tilde{\nu}}, \quad \tilde{\nu} = 8,079 V_v.$$

¹ BIRGE, *Rev. Modern Physics*, vol. 1, p. 1 (1929).

² *Ibid.*, *Phys. Rev.*, vol. 49, p. 204 (1936).

³ *Ibid.*, vol. 54, p. 972 (1938).

⁴ DUNNINGTON, *Rev. Modern Physics*, vol. 11, p. 65 (1939).

INDEX

Boldface numbers indicate references of special importance.

A

A coefficient, Einstein's, 304, 307
 Aberration of light, 29
 Abraham, 60*n*.
 Absorption of light, 247, 305
 coefficient of, 163, 519, 522
 by molecules, 420, 428, 429
 by neon, 247
 by sodium, 248
 Absorption limits, 503, 506, 508
 tables of, 502, 527
 Absorptivity, 158, 162, 163
 Ader, 468*n*.
 Agostino, 617*n*., 618*n*.
 Air pump, invention of, 20
 Alhazen, 10
 Alkali metals, doublet separations, 365
 spectra, 345-353, 362-366
 term energies of, 350
 Allen, 456
 Allison, 487, 535
 Alpha rays or particles, 567, 568-573
 range of, 571-573
 scattering of, 220-224, 591
 theory of, 597-598
 Amaldi, 615*n*., 617*n*., 618
 Ampère, 38
 Anaxagoras, 5
 Anderson, 610, 675, 690
 Andrews, 522*n*.
 Ångström, 207, 210
 Angstrom ($=10^{-8}$ cm.), 207
 Angular momentum, of atom, 339, 355, 412
 with one electron, 299, 302
 in Bohr's theory, 228-229, 302
 of molecule, 419, 430, 431
 of nucleus, 411
 orbital and spin, 309, 342, 355
 principle of, 28
 as vector, 299

Arago, 36, 41
 Archimedes, 9
 Argon, central field for, 314
 Aristarchus, 9
 Aristotle, 6, 14
 Aston, 557*n*., 559, 560, 564*n*., 565*n*., 590, 699
 Atom, Bohr theory of, 226-233, 256
 early views of, 217
 nuclear theory of, 226-233, 256
 Thomson's theory of, 219
 wave-mechanical theory of, for complex atoms, 311-320
 for one-electron atoms, 297-303
 Atomic heat, 442
 Atomic mass unit (a.m.u.), 564, 565, 589, 708
 Atomic numbers, 224, 320, (table) 699
 Atomic weights, 565, (table) 699
 Auger, 549, 679, 696
 Auger effect, 548, (in nucleus) 582
 Average weight of energy levels, 362
 Avogadro's number, 97, 182, 708

B

B coefficient, Einstein's, 305
 Babba, 682*n*., 685*n*.
 Bacher, 254*n*., 619*n*., 621
 Back, 395, 405, 413, 414
 Bäcklin, 96, 544*n*.
 Bacon, Roger, 10
 Bainbridge, 562
 Baker, 621
 Balard, 36
 Balmer, 210, 212
 Balmer series, 210, 211, 214, 370, 371
 Bands, 417, 423, (photographs) 436
 branches and systems of, 433-435
 electronic, 418, 433-437
 head of, 417, 434
 from isotopes, 437

- Bands, rotation, 417, 419–421, 426, 428, 435
 vibration-rotation, 418, 423–429
 effect of temperature on, 428
 in relation to rotation bands, 426
 Barkas, 588*n.*, 699
 Barkla, 477, 480, 481, 483–486
 Barnes, 550*n.*
 Barometer, invention of, 20
 Bartholinus, 36
 Bearden, 494, 542*n.*, 544, 554*n.*
 Beattie, 455*n.*
 Becker, J. A., 535
 Becker, R., 60*n.*
 Becquerel, 567
 Bennett, 648*n.*, 661
 Benoist, 207, 473
 Bernouilli, 28
 Bertin-Sans, 475
 Beta rays or particles (β^\pm), 567
 origin of, 582, 587, 598
 primary and secondary, 579, 582
 spectra of, 582
 Bethe, 279, 591*n.*, 612*n.*, 619*n.*, 626*n.*, 683
 Bierens de Hahn, 205*n.*
 Biot, 36, 38
 Birge, 64*n.*, 97*n.*, 205*n.*, 235*n.*
 Black, 29
 Black body, 160
 radiation, discussion of, 161, 162, 166–170, 173–179, 189–191
 Planck's law, 191, 200–205, 306, 307
 Rayleigh-Jeans formula, 189, 202
 Wien's formula, 190, 202
 Blackett, 668, 695*n.*
 Bless, 535
 Bloch, 536*n.*, 547*n.*, 682*n.*
 Bohr, 258, 641
 on hydrogen, 226, 227, 229, 230, 234, 236, 237, 256
 Bollman, 517*n.*
 Boltzmann distribution law, 255
 Boltzmann factor, 192
 Boltzmann's constant, 182, 708
 Bonner, 634, 638
 Booth, 641*n.*
 Born, 125*n.*, 261, 456
 Bothe, 535, 538*n.*, 606, 666, 696
 Bowen, 663
 Bradley, 29
 Bragg, 488, 488*n.*, 491, 492, 495, 522
 Bragg planes, 488
 Bragg's law, 490, 539
 Brahe, Tycho, 17, 29
 Breadth of spectral lines, 414, 550
 Breit, 596*n.*
 Brewster, 14*n.*
 Brickwedde, 411*n.*
 Brode, 673
 Brown, 183
 Brownian motion, 183
 Brubaker, 634
 Brüche, 659*n.*
 Buguet, 482
 Buisson, 415
 Bunsen, 36
 Bürmeister, 424
 Bursts of ionization, 676
- C
- Cabeo, 20
 Cady, 555*n.*
 Calcite crystals, 493, 494
 Cameron, 644, 645, 646
 Canal rays, 406
 Cario, 250, 251
 Carlisle, 37
 Carlson, 685*n.*
 Carmichael, 688
 Carnot, 33
 Cathode rays, 92
 Cavendish, 30, 31
 Cavendish Laboratory, 48
 Celsius scale, 29
 Chadwick, 568*n.*, 600, 606, 607–609, 631–634
 Characteristic frequencies, 447–449, 457
 functions, values, 296
 temperatures, 452
 X-rays, 495
 Charge cloud, 300, 313, 329
 Charged particles, energy loss by, 681–685, 690–691
 ionization by, 671–674, 681
 measurement of charge, energy, mass, 669–670, (plot) 671
 motion in magnetic field, 526, 670, (plot) 671
 of dipole, 651–655
 of earth, 656–661

- Charged particles, observations on individual, 579, 583-586, 665-669
 Chemical combination, basis of, 320, 329-332
 of certain elements, 320-337
 Childs, 437*n*.
 Chlorine, specific heat, 459, 460, 467
 Clark, 488*n*.
 Clausius, 34, 48
 Clay, 648, 650, 661
 Cleaves, 427*n*.
 Cloud chamber, 584
 Coblenz, 201, 205
 Cockroft, 602, 603-604
 Coincidences, method of, 666
 Collisions of the second kind, 248
 Color vision, 48
 Compton, A. H., 481, 487*n*., 531, 535, 542, 544
 on cosmic rays, 643*n*., 648, 650, 661, 662
 Compton, K. T., 100
 Compton effect, 531-539, 612, (fig.) 680
 Condon, 313*n*.
 Configuration, 375, 376
 Constable, 607*n*., 631-634
 Contraction, in space and time, 136
 Lorentz-Fitzgerald, 130
 Coolidge, 472*n*.
 Cooksey, 623*n*.
 Cooper, J. N., 550*n*.
 Cooper, L., 8*n*.
 Copernican system, 11, 16, 18, 29
 Copernicus, 11
 Corpuscles, 95
 Correspondence principle, 342
 Corson, 673
 Cosmic rays, 643-697
 absorption, in air, 644
 of particles, 666, 668, 678, 691, 695
 in water, 645, 648-649
 discovery of, 643
 ionization due to, 644, 648
 bursts of, 676
 measurement of, 646
 variation, with altitude, 649, 688
 with latitude, 650, 661-664
 with temperature, 662, 694
 with time, 697
 nature and origin of, 650, 697
 particles, 665-669, 689-696
 Cosmic rays, penetrating and soft components of, 692, 696
 showers of, 674, 676-688, 694, 696
 Coster, 546*n*., 548
 Coulomb, 30, 31
 Counters, Geiger or valve, 584
 telescope of, 668
 Coupling, *jj*, 380
 LS or Russell-Saunders, 345, 353-358, 377, 383
 in molecules, (*Λ S*) 430, (*Ω*) 432
 Crane, 613*n*.
 Crussard, 692
 Crystal diffraction grating, 486
 Curie, Marie, 567
 Curie, Pierre, 567
 Curie-Joliot, 606, 611, 613
 Cuykendall, 522*n*.
 Cyclic interchange of variables, 58
 Cyclotron, 623
 Czerny, 420
- D
- D, 343, 348
d, 317
 Darwin, 294, 309, 367
 Davis, 252
 Davisson, 273, 275, 277, 278, 281
 Davy, 33, 40
 De Broglie, 260, 261, 262, 263, 267, 269, 284
 Debye, 450
 Dee, 603
 Degeneracy, 298, 315, 353-355, 384
 nuclear, 411
 Degrees of freedom, 179
 in an enclosure, 184
 energy of, 180-183
 Democritus, 6
 Dempster, 561*n*., 562, 699
 Descartes, 22
 Deuterium, 411
 Deuteron, 574, 589, 609
 Dewar, 209
 Diamagnetism, 45, 387, 389
 Dieke, 440
 Dirac, 261, 309
 Dispersion, 22, 64
 Displacement currents, 48, 50, 51-54

Displacement law, 175, 204
 Dissociation, heat of, 422, 427
 Distribution law, Boltzmann, 255
 Doan, 542, 544
 Doppler, 36
 Doppler effect (moving mirror), 172
 on spectral lines, 414
 Dorsey, 64
 Dow, 572*n*.
 Drinkwater, 371*n*.
 Druyvesteyn, 546*n*., 547*n*.
 Duane, 513*n*., 535
 Du Bridge, 105*n*., 121
 Du Fay, 30
 Dulong and Petit, law of, 442
 Du Mond, 517*n*., 536*n*.
 Dunlap, 617*n*.
 Dunning, 641*n*., 618*n*.
 Dunnington, 96, 96*n*., 517
 Dushman, 286*n*.
 Dymond, 688

E

Eckart, 279
 Edison, 109
 Edlén, 241
 Edwards, 427*n*.
 Eigenvalues (eigenfunctions), 296
 Eighteenth century, physics in, 28-32
 Einstein, 103, 183, 306, 445, 448
 on relativity, 122, 131, 132, 150, 151, 151*n*., 152
 Einstein frequency condition, 227
 Electricity, discoveries in, 30-32, 36-38
 Electrolysis, 37, 43
 Electromagnetic field, 48-61
 energy in, 59
 momentum in, 59
 theory of, 33, 48, 49, 50, 51-61
 Electron, 81, 95, 610
 charge, 95, 97, 708
 discovery of, 92
 e/m 91, 94, 97, 708
 energy loss by, 681-685
 mass, 97, 240, 588, 708
 motion in magnetic field, 526, 670, (plot) 671
 range in air or Pb, 684
 rest energy, 58, 708
 Electrons in complex atoms, 313-317, 375-376, 379, 381, 477, (table) 705
 transitions of, 338, 345, 369, 376, 505
 creation of, 587, 611, 612
 equivalent, 379
 free, 113, 118-121
 in molecules, 429-432
 recoil, 538, 579
 Electron-volt, ev, 237, 708
 relation to wave number, 246, 708
 Ellis, 568*n*., 577, 578, 580, 581
 Elster, 84, 85, 100, 109, 643
 Emissive power, 155, 156, 162, 170
 Emmissivity, 163
 Empedocles, 5
 Endoergic and exoergic, 627
 Energy, conservation of, 33, 34, 44, 583
 per degree of freedom, 180-183
 density of electromagnetic, 59, 65, 158, 160
 in isothermal enclosure, 161, 170
 inertia of, 145, 589, 606
 loss by charged particles, 681-685, 690, 691
 nuclear sources of, 580, 642
 of oscillators in equilibrium, 196
 relativistic, 143, 144, 670
 rest, 145, 589
 of thermal agitation, 194, 195
 Energy levels, 197, 198, 296
 in cm^{-1} , 237
 of harmonic oscillator, 297
 in magnetic field, 387, 392, 397
 molecular, 419, 422, 423, 430, 432
 nuclear (*see* Nucleus)
 of one-electron atom, 298, 368
 X-ray, 504-511, (resonance) 551-552
 Estermann, 284
 Epstein, 407, 658
 Equipartition of energy, 180-184
 Equivalent electrons, 379, 383
 Ether, 140
 Euler, 28
 ev (=electron-volt), 237, 708
 Everblud, 596*n*.
 Exchange effect, 330, 374
 Excitation of atoms, 243-249
 Excitation potentials, 247
 methods of observing, 251

Exclusion principle, 315-316
Expansion chamber, 584

F

F, 343, 348
f, 317
Fabry, 207, 415
Fahie, 14*n*.
Fahrenheit, 29
Falling bodies, laws of, 8, 14
Fankuchen (fig.), 282
Faraday, 38-45, 86, 87
Faraday (unit of charge), 86, 708
Faraday effect, 44
Farkas, 467*n*.
Fast ions, production of, 621
Feather, 609
Fermat's principle, 266
Fermi, 583, 615*n*., 616-618, 638, 639
Fermi-Dirac gas, 114
Ferromagnetism, 335, 387
Field, Coulomb, 313
 modified central, 313
Field currents, 113
Fine structure, 25
 in alkali-type spectra, 362
 in molecular spectra, 432
 for one-electron atoms, 366-372
Fine-structure constant, 367, 708
Fink, 618*n*.
Fitzgerald, 130, 137
Fizeau, 125
Flügge, 565*n*., 699
Fluorescence, 155, 249, 438, 439, 482
Footc, 244, 244*n*.
Force, 28, 143, 144
Försterling, 457*n*.
Fortrat diagram, 434
Fowler, A., 353*n*.
Fowler, R. H., 456*n*.
Frames of reference, 123
 inertial or Galilean, 131, 152
Franck, 247, 250*n*.
Franklin, 30
Fraunhofer, 36
Frenkel, 359*n*.
Frequency (in spectroscopy), 207
Fresnel, 35, 36, 125, 126
Friedrich, 487
Frisch, 284, 639

Froman, 676*n*.
Foucault, 36

G

Galileo, 12-17, 29
Galvani, 37
Gamma rays, 567
 internal conversion of, 579
 origin of, 581, 625, 628-631
 scattering of, 529-531, 537
 spectra, 577
Gas constant, 182, 708
Geiger, 221, 222, 223, 224, 538*n*., 569
Geitel, 84, 85, 100, 109, 643
Gellibrand, 20
Gerlach, 407
Geomagnetic (defined), 656, 659
Germer, 273, 275, 277, 278, 281
Giauque, 437
Gilbert, 20
Gill, 676
Glancing angle, 489
Glazebrook, 46*n*
Gockel, 644
Goldhaber, 609
Goucher, 252
Goudsmit, 254*n*., 309
Gouy, 475
Gram-atom, 86
Gram-ion, 86
Gram-molecule, 86
Gravitation, 20, 21, 26, 150, 152
Gray, L. H., 613
Gray, S., 30
Green, G., 37
Green, G. K., 635
Greenhill, 8
Greinacher, 602*n*.
Grosse, 641*n*.
Ground state = normal state, 231,
 (table) 705
Guericke, 20

H

Hafstad, 630*n*.
Haga, 475
Hagenow, 481
Hahn, 565*n*., 639, 699
Hallwachs, 84

- Hamilton, 32, 262
 Harper, 667*n*.
 Hart, 21*n*.
 Hartley, 209
 Hartley's law, 209
 Hartree, 314
 Haynes, 649
 Heat, caloric theory of, 29, 33, 34
 discoveries in, 29, 33-34
 kinetic theory of, 29, 32
 mechanical equivalent, 34
 Heath, 5*n*.
 Heisenberg, 261, 293, 437
 Heitler, 531*n*., 683, 685*n*.
 Helium, isotope He^3 , 615-616, 638
 mass and binding energy, 588-589
 production of, 570
 spectrum of ionized, 239, 371
 Helmholtz, 34, 50
 Henderson, 605
 Henry, 45
 Herb, 629*n*.
 Hertz, G., 247
 Hertz, H., 50, 83
 Herzberg, 384*n*., 417*n*., 427, 430*n*.
 Herzog, 694
 Hess, 644
 Hevesy, 568*n*.
 High potentials, production of, 602, 622
 High-speed ions, production of, 621
 Hilberry, 695
 Hjalmar, 539
 Hoag, 695
 Hoffmann, 676
 Holes in subshells, 331, 508
 Holloway, 572, 584
 Homonuclear molecules, 429, 441
 Homopolar compounds, 330
 Hooke, 28
 Houstoun, 270*n*.
 Howe, 544*n*.
 Hudson, C. M., 629*n*.
 Hudson, J. C., 502*n*.
 Hudspeth, 617*n*.
 Huggins, 210
 Hughes, 105*n*.
 Huizinga, 546*n*.
 Hull, A. W., 525*n*.
 Hull, G. F., 69
 Hunt, 513*n*.
 Hurmuzescu, 473
 Huygens, 28
 Hydrochloric acid, specific heat of, 468
 spectrum of, 420, 421, 424, 427-429, 432
 Hydrogen, Bohr's theory of, 233-235
 energy levels, 237, (table) 238, (diagram) 242
 heavy, 411, 566
 orthohydrogen, 466
 parahydrogen, 466
 spectrum of, (figs.) 211, 237, 241, 371
 fine structure, 366-371
 isotope structure, 410
 Hyperfine structure, 409-414

 I
 I, (nucleus) 411
 Idei, 512*n*.
 Imbert, 475
 Imes, 424
 Impenetrability of molecules, 330
 Indeterminacy principle, 292-294, 339
 Induced emission, 306
 Inert gases, 322, 384, 458
 Intensity of electromagnetic waves, 66
 Intercombination lines or series, 216
 Interference of light, 35
 Ionization of atoms, 243-255, 499
 by charged particles, 671-674, 685
 multiple, 247, 254, 337, 545
 Ionization chamber, 473, 584
 Ionization energy or potential of atoms, 233, 243, 253, (table) 254, (plot) 353, (table) 705
 Isobars, 595
 Isolelectronic atoms, 338
 Isothermal enclosure, 161
 Isotope structure, 409
 Isotopes, 560, 562-566, (table) 699
 separation of, 565, 637, 641
 Isotopic weight, 563 (table) 699
 Ives, 138*n*.

 J
 Jeans, 179, 184, 390*n*.
 Johnson, 658*n*.
 Johnston, 437
 Joliot, 606, 610, 613

Jones, 522*n*.
Jongen, 648
Joule, 34

K

k, 182, 205, 708
K lines or series, 486, 497
Kaufman, 512*n*.
Kayser, 208, 211
Keesom, 455
Kellogg, 412*n*.
Kelvin, 34
Kennard, 116*n*., 514*n*.
Kepler, 17, 69
Kepler's laws, 18, 19
Kerr, 45*n*.
Kikuchi, 281
Kinetic theory of gases, 48
Kircher, 20, 29
Kirchhoff's law, 163, 306
Kirkpatrick, 518*n*., 536*n*.
Klein, 537*n*.
Klemperer, 613*n*.
Knipping, 487
Knox, 36
Kolhörster, 644, 645, 666
Korff, 696
Kossel, 553*n*.
Kramers, 437
Kretschmar, 528*n*.
Krishnan, 437
Kronig, 548
Kruger, 635
Kuhn, 467*n*.
Kuipers, 546*n*.
Kulenkampff, 518
Kurlbaum, 191
Kussmann, 205*n*.

L

L lines or series, 486, 497
Ladenburg, 100, 205
Lagrange, 28
Larsson, 540
Landé, 398
Landé *g* or splitting factor, 397-403
 table, 401
Landé's interval rule, 361
Langmuir, 258

Laplace, 37
Larmor, 390*n*.
Larmor precession, 390
Laue, 486
Laue spots, 491
Lauritsen, 613*n*.
Lawrence, 623, 625
Lead, spectrum of, 382
 X-rays from, 520, 521
Least action, 266, 268
Lebedev, 69
Leibnitz, 28
Lemaitre, 658, 660
Lenard, 85, 98, 100, 251
Lenses, discoveries concerning, 20, 22,
 24
Leprince-Ringuet, 692
Leu, 409
Lewis, G. N., 258, 566*n*.
Lewis, W. M. C., 456*n*.
Libration, 257
Light, corpuscular theory of, 24, 32, 33,
 108, 109, 260
 discoveries in, 29, 34-36
 electromagnetic theory of, 49
 velocity of, 30, 36, 63, 64, 134, 138
 wave theory of, 24, 32-36, 109, 260
Lightning, 31
Lindemann, 448, 450
Line breadth, 370
Lipperhey, 14
Liveing, 209
Livingood, 626*n*.
Livingston, 572, 584, 605*n*., 623, 626*n*.
Locher, 668
Lodge, 14*n*.
Loeb, 192*n*.
Lorentz, 50, 55, 87, 88, 95, 126
 in connection with relativity, 130,
 135, 137
Lorentz transformation, 134
Lorentz unit, 394
Lorenz, 87
Loria, 250*n*.
Lummer, 169, 177, 191
Lyman series, 236

M

MacDonald, 566*n*.
Mach, 28*n*., 122

- Magnetic moment, of atom, 397, 409
 of one-electron atom, 392
 due to orbital motion, 384
 due to spin, 309, 385
 of earth, 658, 659
 of nucleus, 409, 411, (table) 414
 Magnetism, 10, 20
 Magneton, 392
 nuclear, 412
 Maltezos, 475
 Mann, 128*n*.
 Marsden, 222, 223, 224
 Mass, atomic unit of, 564, 565, 589, 708
 electromagnetic, 80
 of energy, 61, 68, 145, 589, 606
 inertial, (defined) 142
 reduced, 232
 rest, 143
 variation of, 140
 Mass defect, 590
 Mass numbers, 563, (table) 699
 Mass spectrograph, 559–563
 Masses of atoms, 556, 588, (table) 699
 spectrum of, (photograph) 563
 Mattauch, 565*n*., 699
 Matter, nature of, 86
 Matrix component, 307
 Maxwell, 46–49, 50, 52, 65
 Maxwell's equations, 54
 Maxwell's law, 192
 Mayer, J. E., 192*n*.
 Mayer, M. G., 192*n*.
 Mayer, R. J., 34
 Mean life, 304
 Mechanics, discoveries in, 20, 27, 28, 32
 and geometrical optics, 262
 relativistic, 140
 Mecke, 437*n*.
 Meggers, 244
 Meitner, 639
 Mercury, spectrum of, 250, 377–379, 401
 Mersenne, 20
 Mesotrons (mesons), as cause of show-
 ers, 678, 696
 decay of, 694
 discovery of, 689
 energy loss by, 682
 mass of, 693
 origin of, 693
 showers of, 696
 Metals, electrons in, 113–121
 Metastable levels, 378
 Michelson, 64, 126, 128, 207, 370, 409, 416
 Michelson-Morley experiment, 126, 137
 Micron, 207
 Microscope, electron, 266
 gamma-ray, 294
 Miller, 128*n*.
 Millikan, 95, 96, 103, 183*n*.
 on cosmic rays, 643*n*., 644–649, 661, 663, 690
 Mitchell, 618*n*.
 Mohler, 244, 244*n*.
 Molecules, average energy of, 194, 195
 diatomic, 180, (spectra) 419–437, 459, 464
 velocity distribution, 192
 wave-mechanical theory of, 421, 429
 Møller, 682*n*.
 Momentum, electromagnetic, 59, 68
 indeterminacy of, 292–294
 relativistic, 142, 143, 145
 Montgomery, 679*n*.
 Morgan, 678*n*.
 Morley, 126, 128, 370
 Moseley, 320, 496, 497, 498
 Moseley's law, 496
 Mott-Smith, 668
 Multiple quantum states, 255
 Multiplets, *jj*, 382
 LS or Russell-Saunders, 355, 358, 383
 of one-electron atoms, 366–369
 of spectral lines, 357
 Multiplicity of a term, 356
 Murphy, 411

N

- N*₀, 86
 Nähring, 543
 Neddermeyer, 675, 690
 Negatron, 610
 Neher, 649, 661, 663, 667*n*.
 Neon spectrum, 247, 248
 Nernst, 450, 455*n*., 457
 Neuert, 638
 Neutrino, 583
 Neutrons, 587, 588, 615–621, 696
 capture of, 619
 counting of, 616

- Neutrons, discovery of, 606
 forces on, 593, 596
 mass of, 610
 in nuclear reactions, 606, 614, 615,
 617, 618, 626-629, 631, 636-638,
 642
 scattering of, 619
 slow and thermal, 618
 sources of, 615, 625-629, 631, 634, 642
 velocity of, 616
- Newton, 21-28, 30, 36, 69, 132
- Nichols, 69, 448
- Nicholson, 37
- Nicholson, J. W., 227
- Nickel crystal, 273, 280
- Nielsen, 678*n*.
- Nier, 564*n*., 641*n*.
- Nishina, 537*n*.
- Noncombining terms, 357
- Normal state, 231, (table) 705
- Normalization to unity, 291
- Northrop, 622*n*.
- Nuclear reactions, 601-642
 alternative bombardments, 637
 results, 631, 636
 determination of mass by, 637
 due to neutrons, 628-629
 (See also Neutrons)
 emission of γ -rays, 628-631, 635, 637
 endoergic or exoergic, 627
 energy relations, 603-606, 627, 631
 635, 640, 642
 isotopic product, 628
 many particles produced, 628
 neutron groups emitted, 634
 notation for, 601-602, 626
 proton groups emitted, 629, 631
 with radioactive product, 626, 636-
 637
 P^{30} , 614, 636
 N^{13} , 614, 631
 S^{31} , Al^{196} , U^{230} , 628
 Na^{22} , 631
 Al^{28} , 636
 Na^{24} , 628, 636
 Mg^{27} , 636
 As^{76} , 637
 Ag^{106} , 637
 $Z = 43$, 637
 resonance levels, 631, 633
- Nuclear reactions, types of, 625-626
 $Al^{27} + \alpha \rightarrow P^{30} + n$, 614
 $B^{11} + H^1 \rightarrow Be^8 + \alpha$ or 3α , 605
 $B^{10} + n \rightarrow Li^7 + \alpha$, 616
 $B^{10} + \alpha \rightarrow N^{13} + n$, 614
 $Be^9 + H^1 \rightarrow Be^8 + H^2$, 625
 $Be^9 + H^2 \rightarrow B^{10} + n$, 615, 634
 $Be^9 + \alpha \rightarrow C^{12} + n$, 608
 $Be^9 + \alpha \rightarrow N^{14} + n$, 608
 $F^{19} + H^1 \rightarrow O^{16} + \alpha$, 605
 $F^{19} + \alpha \rightarrow Na^{22} + n$, 631
 $F^{19} + \alpha \rightarrow Ne^{22} + H^1$, 631
 $H^1 + n \rightarrow H^2$, 618
 $H^2 + H^2 \rightarrow He^3 + n$, 615
 $Li^7 + H^1 \rightarrow 2He^4$, 603-604, 629
 $Li^6 + H^2 \rightarrow Li^7 + H^1$, 629
 $Li^6 + n \rightarrow H^3 + \alpha$, 616
 $N^{14} + \alpha \rightarrow O^{17} + H^1$, 601
- Nucleus, 222, 556-642
 angular momentum (or spin), 411
 constituents of, 586
 energy from, 640, 642
 energy levels, 577, 580, 593, 632
 of B^{10} , 635
 of Ne^{22} , 633
 energy of, 589, 590-591, 593, 594
 fission of, 638
 isomeric, 628
 magnetic moment of, 409, 411, 414
 motion of, 232, 320, 431
 radius of, 592
 structure of, 591-597, 641
 theory of, 591-599
 transformation of, 574-577, 599-606,
 689
 (See also Nuclear reactions)
- O
- Occhialini, 668
- Octets of elements, 324
- Oersted, 38
- One-electron atoms, 226-241
 transitions of, 227, 234
- Operators, 360
- Oppenheimer, 685*n*.
- Orbits, elliptical, 256, 302
 frequency of revolution, 236
 in one-electron atom, 231-233, 302
- Oscillators, harmonic, energy levels, 297
 radiation by, 79

Oscillators, in thermal equilibrium, 191–197, 199
 wave-mechanical theory of, 296–297
 Otis, 646*n*.

P

P, 343, 348
p, 317
 Packing fractions, 590, (table) 699
 Pair production, 611, 680
 Paneth, 568*n*.
 Paramagnetism, 387
 Parratt, 512*n*., 546*n*., 550*n*., 553*n*.
 Particles, 290, 294
 in nucleus, 586, 591–593
 Pascal, 20
 Paschen, 371, 395, 405
 Paschen-Back effect, 403
 Paschen series, 236
 Pauli, 315, 410
 Pauling, 296*n*., 332*n*.
 Pegram, 618*n*.
 Peierls, 591*n*.
 Pendulum, 14, 28
 Peregrinus, 10
 Period (of decay), 574
 Periodic system, 320–337
 Bohr's table, 321, 597
 standard table, 704
 Perlow, 638
 Perot, 207
 Perrin, 92
 Perturbation theory, 319, 354
 approximation, first-order, 319
 zero-order, 314–318, 380–382
 Pfotzer, 688
 Phosphorescence, 155
 Photoelectric effect, 83–121
 and Compton effect, 537, 612
 discovery, 83
 due to γ -rays, 578–579, (fig.) 680
 due to X-rays, 523–528, 537, 552
 electrons, energy of, 100–102
 source of, 105, 117–121
 nature of, 85, 97, 106, 108, 111–121
 variation with frequency, 103, 119
 with intensity, 100
 with potential, 98, 101
 various properties of, 104, 121
 X-ray, 484–486, 504

Picard, 27
 Pierce, 522*n*.
 Plain, 629*n*.
 Planck, 69*n*., 190, 191, 197, 198
 Planck's constant, h , 197, 205, 708
 Planck's radiation law, 200–205
 Pohl, 475
 Point charge, uniformly moving, 70, 148
 accelerated, 73–80, 150
 Poisson, 37
 Polarization of light, 35, 36, 62
 Pollard, 607*n*., 699
 Pontecorvo, 618*n*.
 Positive rays, 556
 Positrons, 568, 598, 610–613, (fig.) 674
 Potential barrier (nuclear), 598
 Poynting's vector, 59
 Preston, 400
 Preston's rule, 400
 Pringsheim, 169, 177, 191
 Probability amplitude, 291, 311
 Probability for momentum, 292
 of transition, 304, 307
 wave-mechanical, 291, 300
 Protons, due to bombardment, 600, 626, 627, 629, 631, 636
 energy loss by, 681–684
 forces on, 593, 596
 mass of, 588
 measurement of energy, 671
 range of, 573
 scattering of, 596
 Prout, 217
 Ptolemy, 9
 Pythagoras, 5

Q

Q (nuclear heat of reaction), 604–606, 627
 Quadrupole X-ray lines, 512
 Quantum defect, 352
 Quantum numbers, for atoms, 298, 300, 310, 315, 317
 $n \lambda \mu$, 298, 350
 $n l j m$, 381, 507
 $J M$, 340, 353, 355, 376, 381, 396, 405
 $L \Lambda S \Sigma$, 343, 344, 405
 $L S J M$, 355, 376

- Quantum numbers, for atoms, *I*, 411
 F M_F, 412
 effective principal, 352
 of electrons in atoms, 374–377, 381,
 (table) 705
 in *K* and *L* shells, 318
 for molecules, (*J*) 419, (*v*) 422,
 (*A*) 430–431, (*S*) 431, (*K*) 431,
 (*J M*) 432
 of X-ray levels, 508–510
 Quantum of energy, 198, 200
 Quantum states, 198, 199, 227, 255,
 294, 296
 due to spin, 309
 in external field, 384
 of harmonic oscillator, 197, 297
 of ions, (X-ray) 499
 in magnetic field, 389–390, 391, 398
 of many-electron atom, 312, 314
 of molecules, 429
 diatomic, 419, 422, 423, 431
 multiple, 255
 normal, 231, (table) 708
 nuclear, 593
 of one-electron atom, 230, 297, 298
 Quantum theory, birth of, 190, 197–200
 (See also Wave mechanics)
- R
- R*, *R_∞* (see Rydberg constant)
 Rabi, 409, 412
 Radiation, absorption and emission of,
 304
 by charged particles, 76, 683–685
 dipole, 308
 by harmonic oscillator, 79
 quadrupole, 309
 density of, 158
 in electromagnetic field, 59
 intensity of, 66, 156, 157, 159
 pressure due to, 69, 164, 170
 resonance, 248
 (See also Black-body radiation)
 Radioactivity, 566–583
 definition of, 582
 discovery of, 566
 explanation of, 597
 induced, 613–615, 625, 628(*d*)
 nuclear transformations in, 574–577
 Raman, 437
 Raman effect, 437
 Ramberg, 547*n.*, 550*n.*
 Ramsauer, 282*n.*
 Rare earths, 335
 Rasetti, 568*n.*, 584*n.*, 617*n.*, 618*n.*
 Rayleigh, 179, 184, 415, 437
 Rayleigh (R. J. Strutt), 249
 Rayleigh-Jeans formula, 189, 201
 Reflection from moving mirror, 170
 Reflectivity, 158
 Refraction, 9, 10, 20, 22, 64
 of X-rays, 475, 539–544
 Regener, 646*n.*, 648
 Reich, 572*n.*
 Reimann, 110*n.*
 Relativity, 122–153
 and electromagnetism, 147
 mechanics, 140
 properties of light, 124–130
 wave mechanics, 309, 358
 general theory of, 150
 Newtonian, 122, 134, 139
 special theory of, 130, 131
 Residual rays, 447
 Resonance potentials, 253
 Resonance radiation, 248, 250
 Richardson, 100, 110*n.*, 371
 Richardson's equation, 110
 Richtmyer, 100, 521, 522, 545*n.*, 547*n.*,
 550*n.*
 Ritz, 216
 Ritz combination principle, 216
 Roberts, 630*n.*
 Robinson, 524, 527
 Roentgen, 469–471, 484
 Rojansky, 296*n.*
 Römer, 30
 Rosa, 64
 Rosenberger, 5*n.*
 Ross, 535
 Rossi, 666, 667, 668, 676, 695
 Royal Institution, 33, 40
 Ruark, 256*n.*
 Rubens, 191, 447, 448, 457
 Rumbaugh, 630*n.*
 Rumford, 33
 Runge, 211
 Rupp, 282, 283
 Rutherford, 568, 568*n.*, 569, 570, 599,
 600, 608
 on scattering of α 's, 221–224

Rydberg, 211–217, 322
 Rydberg constant, 214, 215, 235, 241,
 411, 708
 for one-electron atom, 240, 368
 Rydberg formula, 214, 216, 352, 362
 Rydberg-Schuster law, 215, 243

S

S, 343, 348
s, 317
 Sadler, 477*n.*, 483
 Saha, 631*n.*, 633
 Savart, 38
 Scattering, of α particles, 220–224, 591
 by Compton effect, 531–539
 of γ rays, 529–531, 537
 of light, 437
 of matter waves, 290
 multiple and single, 221
 of neutrons, 619
 of protons, 596
 of X-rays, 475–482, 528–539
 angular distribution, 530
 plot, 529
 Schein, 676, 693*n.*
 Scheiner, 20
 Schrödinger, 260, 261, 285, 294, 454, 456
 Schuster, 215
 Screening of nucleus, 499
 Seaborg, 626*n.*
 Segré, 617*n.*, 618*n.*
 Selection rules, in atoms, 308, 339, 376
 l, 346
 j, 369, 381
 λ μ , 393
 J M, 341, 405
 L A, *S Z*, 344, 405
 for molecules, *J*, 419, 424, 429, 433
 K, 431, 433
 A S, 431
 for Raman lines, 439
 for X-ray lines, 509, 512
 Serber, 687
 Series (spectral), discovery of, 210–212
 fundamental, 348
 interrelations between, 212–217
 principal, sharp, diffuse, 212, 347–349
 relation to terms, 216
 Shaw, 542*n.*
 Shells, 317, 318, (table) 336, 508

Shimizu, 584
 Shortley, 313*n.*
 Siegbahn, 494, 502
 Silberstein, 135*n.*
 Simon, 535
 Simultaneity, 132
 Skinner, 550*n.*, 580
 Skobelzyn, 665, 674
 Slater, 332*n.*
 Smekal, 438
 Smythe, 31*n.*
 Snell, 20
 Snyder, 685*n.*
 Sodium, energy levels, (diagram) 349,
 (list) 363
 quantum defects, 352
 spectrum, 212, 241, 248, 348–352,
 363, 364
 excitation of, 245, 246
 Zeeman effect for, 402–404
 Sodium chloride, 492
 Sommerfeld, 114, 256, 256*n.*, 367, 372,
 516
 Sound, discoveries in, 20
 Space quantization, 300
 Space-time, 133
 Specific heats, 442–468
 of gases, classical theory, 457–460
 as observed, 465
 quantum theory, 460–465
 of solids, 442–456
 classical theory, 444
 Debye's theory, 450–456
 Einstein's theory, 445
 mixed solids, 456
 variation with temperature, 442–
 444, 453–456
 Spectra, alkali type of, 345–353, 362–
 366
 arc and spark, 209
 from ionized atoms, 337
 from many-electron atoms, 383
 mode of origin of, 337
 molecular, 417
 electronic, 418, 433–437
 rotation, 417, 419–421
 vibration-rotation, 418, 421–429
 from one-electron atoms, 233–241,
 369
 two-electron type of, 372
 (See also Series)

Spectral lines, breadth of, 414
 Spin, electron, 309, 330, 385, 431
 special correction due to, 368
 neutron and proton, 596
 nuclear, 411
 Spinks, 427*n*.
 Spin-orbit effect, on *LS* terms, 355,
 358-361, 365
 on magnetic levels, 395
 in molecules, 430, 431, 432
 in one-electron atom, 368
 Spin-spin interaction, 366
 Stark, 406
 Stark effect, 406
 Stationary states, 227, 294, 296, 312
 Stearns, 661, 676*n*.
 Stefan-Boltzmann law, 166-170, 204
 Stenström, 539
 Stern, 284, 407
 Stern-Gerlach experiment, 407
 Stevenson, 693
 Stilwell, 138*n*.
 Stokes' law, 96
 Stoletow, 84
 Stoney, 87, 474
 Störmer, 658
 Strassmann, 639
 Straubel, 483
 Street, 693
 Strutt, 249*n*.
 Subshells, 317, 318, 330, (table) 336,
 382, 508

T

Tarrant, 613
 Taylor, 409
 Telescope, 14, 15, 24
 Terms (spectral), 215, 216, 238
 of alkali metals, 350
 doublet, 362
 jj, 380
 LS, 345, 353-362
 singlet and triplet, 374
 Thales, 5
 Thaxton, 596*n*.
 Thermal radiation, 154
 Thermionic emission, 109-113
 origin of thermions, 117
 Thermodynamics, 33, 34
 Thermometer, invention of, 29

Thibaud, 544*n*.
 Thompson, 38*n*.
 Thomson, G. P., 118*n*., (fig.) 282
 Thomson, J. J., 221, 557, 558, 559
 on electrons, 80, 93, 94, 95, 109
 on X-rays, 218, 471, 476
 Time (in relativity), 122, 132
 Tolman, 192*n*.
 Tomboulion, 550*n*.
 Torricelli, 20
 Townsend, 95
 Tracks, (photograph) 585
 Transition curves, 677
 Transitions, probability of, 304, 307,
 308
 radiationless, 548, 582
 Transmissivity, 158
 Transmutation of atoms, 574-577,
 599-642
 Trautz, 468*n*.
 Turner, 638*n*., 648*n*., 650, 661, 662
 Tuve, 667*n*.

U

Uhlenbeck, 309
 Ulrey, 513*n*.
 Uranium, fission of, 638
 X-ray diagram for, 510
 Urey, 256*n*., 411

V

Valence bonds, 329-332
 Vallarta, 658, 660
 Van Atta, 622*n*.
 Van de Graaff generator, 622
 Van den Ende, 445
 Vector diagrams, 300
 for *J M*, 341
 for *L S J*, 356
 for Zeeman effect, 393, 396, 397
 Velocities, transformation of, 138
 Villari, 472
 Vinci, Leonardo da, 10
 Viscosity of gases, 48
 Vogt, 502*n*.
 Volta, 37

W

Wagner, 518
 Walter, 475

Walton, 602, 603, 604
 Warburton, 522*n*.
 Wave equation, 288
 for harmonic oscillator, 296
 for a one-electron atom, 298
 for a particle, 288, 295
 Wave functions, 296
 of harmonic oscillator, 297
 of one electron, in complex atom, 314
 of one-electron atom, 298, 301
 Wave length, de Broglie (of matter waves), 269-272
 measurement of, 35, 207-209
 Wave mechanics, of many-electron systems, 311-320
 of one particle, 259-310
 relativistic, 309, 358
 Wave number, $\tilde{\nu}$, 207
 in terms of electron-volts, 246, 708
 Wave packets, 291-294, 599
 Waves, electromagnetic, 61-70
 matter (*e.g.*, electron), 259-292
 diffraction of, 273, 281, 283
 geometrical optics of, 262
 physical significance of, 289-292
 refraction of, 263, 278
 wave length of, 269, 271
 phase, 261
 standing, in a box, 187
 Weber, 443
 Weizel, 417*n*.
 Wentzel, 547*n*.
 White, A. D., 16*n*.
 White, H. E., 353*n*.
 Wien, 179, 190
 Wien displacement law, 175, 204
 Wien's formula, 190, 191, 201
 Williams, R. C., 370
 Williams, W. E., 371*n*.
 Wilson, C. T. R., 535, 584*n*., 643, 672
 Wilson, E. B., Jr., 296*n*.
 Wilson, V. C., 693*n*.
 Wilson, W., 672
 Wind, 475
 Winkelmann, 483
 Wollan, 648*n*.
 Wood, 248, 249, 249*n*., 440
 Wooster, 581
 Work function, 111, 118
 Wulff, 414
 Wynn-Williams, 584*n*.

X

X-rays, 469-555
 absorption of, 482-486, 502-504, 519-523, 537
 lines, 550-553
 absorption limits, 503, 506, 508, 552
 tables of, 502, 527
 Auger effect, 548
 characteristic, 483, 495
 continuous spectrum, 512-519
 limiting frequency of, 514
 critical voltages, 503
 diffraction of, 475, 486-491, 537, 544
 discovery and properties of, 469
 efficiency of production of, 519
 emission of, 474, 475, 499, 515
 energy levels for, 504-511
 quantum theory of, 507-512
 fluorescence, 484-486, 504
 frequencies, 497, (table) 501-502
 hardness of, 483
 homogeneous, 483
 intensity of, continuous spectrum, 517
 lines, 512
 measurement of, 472, 517
 involving periphery of atom, 550
 from multiply ionized atoms, 545
 photoelectric effect, 523-528, 537, 552
 polarization of, 480, 536
 pulse theory of, 473-474, 476-480
 reflection of, total, 542
 refraction of, 475, 539-544
 scattering of, 475, 480-482, 528-539
 angular distribution, 530
 classical (Thomson) theory, 476
 coefficient of, 528, (plot) 529
 Compton, 531-539
 second-order (or satellite) lines, 545
 from solids, 553
 spectra, 486, 495-496, 505-507, 512
 origin of, 498
 plots of, 500, 502
 selection rules for, 509, 512
 spectrometer for, 491-495
 target thickness, effect of, 516
 tubes for, 471
 wave lengths, 475, 494, 496, (table) 501-502
 measurement of, 542, 544
 X-unit, 207, 493-494

Y

Young, 34, 48

Z

Zacharias, 412*n*.

Zeeman, 87

Zeeman effect, 87-92, **334-406**

Zeeman effect, anomalous, 384

classical theory of, 87-92, 387-391

in huge field, 391-395

normal triplet, 384, 394, 401, 405

Paschen-Back effect, 403

photographs of, 404

in strong field, 403-409

in weak field, 396-403

Zinc, spectrum of, 209